HUMBOLDT-UNIVERSITÄT ZU BERLIN

Functional characterization of the RNA-binding protein HDLBP

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Ulrike Zinnall
geboren am XX.XX.XXXX
in XXXXXXX

Präsidentin der Humboldt-Universität zu Berlin: Prof. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin: Prof. Bernhard Grimm

Gutachter der Dissertation:
Prof. Markus Landthaler
Dr. Benedikt Beckmann
Prof. Florian Heyd

Tag der mündlichen Prüfung: 18.12.2020

TABLE OF CONTENTS

STATEMENT OF CONTRIBUTIONS

The functional characterization of HDLBP would not have been possible without the help of many people. The project contribution was as follows:
I performed the majority of the experimental work. Miha Milek (Landthaler lab) carried out the computational data analyses and performed the ncRNA-sequencing. This project was supervised by Markus Landthaler and discussed with the whole Landthaler lab. Ouidad and Nouhad Benlasfer (Landthaler lab) helped with the generation of HEK293 stable cell lines (see chapter 1). The mass spectrometry and MaxQuant runs for the BioID experiment (see chapter 6) were carried out by Guido Mastrobuoni from the lab of Stefan Kempa (MDC). The mass spectrometry and MaxQuant runs for the pSILAC experiment (see chapter 8) were performed by Carlos H. Vieira-Vieira from the lab of Matthias Selbach (MDC). Bettina Purfürst (MDC) carried out the electron microscopy imaging (see chapter 7). The settings of the flow cytometer instrument were adjusted with help of Caroline Braeuning (MDC) (see chapter 10). Simon Müller from the lab of Stefan Hüttelmaier (Martin-Luther-University Halle-Wittenberg) carried out phenotyping of A549 cells and mouse xenograft experiments (see chapter 11).

Parts of this thesis (especially the materials and methods section and the figures) will be published as:
"HDLBP discriminates between ER-targeted and cytosolic mRNAs to promote protein synthesis of transmembrane and secreted proteins"
Ulrike Zinnall[+], Miha Milek[+], Carlos H. Vieira-Vieira, Simon Müller, Guido Mastrobuoni, Orsalia-Georgia Hazapis, Igor Minia, Simone del Guidice, Nadine Bley, Stefan Kempa, Stefan Hüttelmaier, Matthias Selbach, Markus Landthaler
[+]These authors contributed equally
[Submitted to Nature Communications, under revision, as of October 2020]

SUMMARY

The secretory pathway is essential for proper cell functioning and starts when mRNAs encoding membrane and secretory proteins are targeted to the endoplasmic reticulum (ER). However, little is known about the contribution of RNA-binding proteins to the recognition, localization and translation of ER-localized mRNAs.

In this work, we characterized the human RNA-binding protein HDLBP by using a variety of biochemical, transcriptomic and proteomic approaches. We identified that HDLBP binds to more than 80% of all ER-localized mRNAs by PAR-CLIP, cell fractionation and RNA-sequencing experiments. Analysis of the HDLBP binding motif showed that it predominantly binds to a CU-containing motif and forms high affinity multivalent interactions primarily in the coding sequence (CDS) of ER-localized mRNAs. In contrast, we identified that cytosolic HDLBP mRNA targets show less HDLBP binding sites randomly distributed between the CDS or 3' untranslated regions. This indicates that ER-localized mRNAs per se differ from cytosol-localized mRNAs in their sequence composition with regard to HDLBP binding sites.

Further PAR-CLIP analysis revealed that HDLBP interacts with RNA components of the signal recognition particle (SRP) and the 40S ribosomal subunit. We identified by BioID experiments proteins in close proximity to HDLBP and thereby confirmed the association of HDLBP with components of the translational apparatus and the SRP.

Functional studies using CRISPR-Cas9 HDLBP knockout (KO) cell lines in combination with ribosome profiling demonstrated that HDLBP promotes the translation of its ER-localized target mRNAs. We validated this finding by pSILAC experiments and detected the corresponding decrease in protein synthesis of proteins encoded by mRNAs that are bound by HDLBP and ER-localized. Furthermore, we determined the effect of HDLBP on secretion and membrane-protein surface expression and found that both are impaired in absence of HDLBP.

Lastly, *in vivo* experiments with nude mice showed that HDLBP KO resulted in a decrease of lung tumor formation highlighting the relevance of HDLBP for tumor progression. Overall, these results demonstrate a general function for HDLBP in the translation of ER-localized mRNAs.

ZUSAMMENFASSUNG

Der Sekretionsweg ist essenziell für die ordnungsgemäße Funktion von Zellen und beginnt, wenn mRNAs, die für Membran- und Sekretionsproteine codieren, an das endoplasmatische Retikulum (ER) gebracht werden. Allerdings ist wenig darüber bekannt, inwiefern RNA-bindende Proteine zur Erkennung, Lokalisierung und Translation von ER lokalisierten mRNAs beitragen.

In dieser Arbeit haben wir das humane RNA-bindende Protein HDLBP mithilfe verschiedener biochemischer, transkriptomischer und proteomischer Methoden charakterisiert. Wir haben durch PAR-CLIP-, Zellfraktionierungs- und RNA-Seqeuenzierexperimente festgestellt, dass HDLBP an mehr als 80% aller ER lokalisierten mRNAs bindet. Analysen zu HDLBPs Bindungsmotiv haben gezeigt, dass HDLBP vorwiegend an ein CU-haltiges Motiv bindet und hoch affine, multivalente Wechselwirkungen mit der codierenden Sequenz (CDS) hauptsächlich von ER lokalisierten mRNAs eingeht. Im Gegensatz dazu haben wir festgestellt, dass zytosolische HDLBP gebundene mRNAs weniger Bindungsstellen aufweisen und dass diese sowohl in der CDS als auch in 3‘ untranslatierten Regionen auftreten. Dies zeigt, dass sich ER lokalisierte mRNAs von Zytosol lokalisierten mRNAs in ihrer Sequenzzusammensetzung hinsichtlich der HDLBP Bindungsstellen unterscheiden.

Weitere Analysen des PAR-CLIP-Experiments ergaben, dass HDLBP mit RNA-Komponenten des Signalerkennungspartikels (SRP) und der 40S ribosomalen Untereinheit interagiert. Durch BioID-Experimente haben wir Proteine in unmittelbarer Nähe zu HDLBP bestimmt und konnten damit die Assoziation von HDLBP mit Komponenten des Translationsapparates und des SRPs bestätigen.

Funktionelle Studien, bei denen wir CRISPR-Cas9 erzeugte HDLBP Knockout (KO) Zelllinien in Kombination mit Ribosomen-Profiling verwendet haben, haben gezeigt, dass HDLBP die Translation von mRNAs fördert, die an HDLBP gebunden und am ER lokalisiert sind. Dies haben wir durch pSILAC-Experimente validiert und die entsprechende Abnahme der Proteinsynthese von Proteinen festgestellt, die von mRNAs codiert werden, die an HDLBP gebunden und am ER lokalisiert sind. Darüber hinaus haben wir die Wirkung von HDLBP auf die Sekretion und die Oberflächenexpression von Membranproteinen untersucht und herausgefunden, dass HDLBPs Abwesenheit beides beeinträchtigt.

Letztlich haben *in vivo* Experimente mit Nacktmäusen ergeben, dass HDLBP KO eine Abnahme der Lungentumorbildung verursacht, was die Relevanz von HDLBP für die Tumorprogression hervorhebt. Insgesamt zeigt unsere Arbeit eine generelle Funktion von HDLBP bei der Translation von ER lokalisierten mRNAs.

INTRODUCTION

The secretory pathway

Secretion of certain proteins to the extracellular space is essential for eukaryotic cells and needed for multiple purposes like cell homeostasis, hormone release and neurotransmission. Constitutive secretion is fundamental for every cell and key factors of the pathway are ubiquitously expressed and highly conserved. In addition, there are specialized secretory cells such as pancreatic beta-cells secreting insulin. Here, the secretion is regulated and triggered by extracellular stimuli. The conventional secretory pathway includes both constitutive and regulated secretion and covers all steps from mRNA translation to the secretion of the mature protein (Viotti, 2016). A schematic overview is shown in Figure I. Moreover, the secretory pathway includes the biosynthesis of membrane proteins which is similar to secretory protein biosynthesis. Only the last step differs because membrane proteins are integrated into the plasma membrane, whereas secretory proteins are secreted. Each step is regulated and quality control pathways are in place to ensure the correct procedure (McCaffrey and Braakman, 2016). Mis-regulation can lead to several disease phenotypes like cancer or neurodegenerative diseases (Wang and Kaufman, 2016).

The secretory pathway starts when mRNAs encoding a signal sequence or transmembrane domain are co-translationally targeted to the endoplasmic reticulum (ER). During translation the nascent peptides are translocated into the ER lumen where they are folded and packed into vesicles. These vesicles are transported to the Golgi apparatus. After several posttranslational modifications the mature proteins are encapsulated into secretory granules, leave the Golgi apparatus and are transported via the cytoskeleton to the plasma membrane. The granules fuse with the plasma membrane and the proteins are secreted or integrated into the plasma membrane.

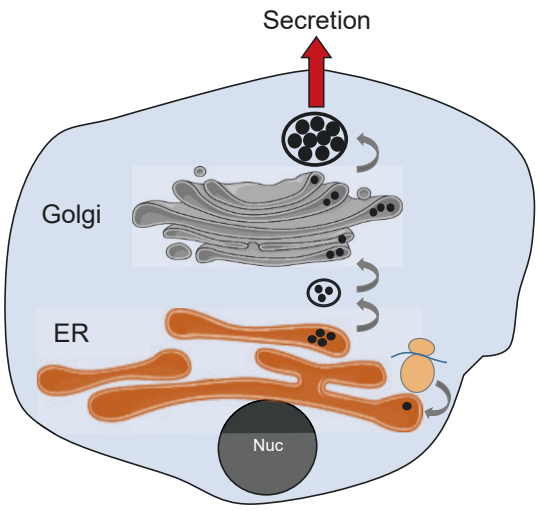

Figure I | Overview of the secretory pathway.

Besides the conventional secretory pathway, alternative routes that do not involve the ER-to-Golgi transport are described as unconventional protein secretion pathways (Rabouille, 2017).
Here, we will focus on the very first steps of conventional secretion where the ER plays a major role.

The ER

The ER consist of a flattened membrane enclosed sacs or tubes known as cisternae. This membrane network can have a surface area that is 25 fold greater than the membrane surrounding it. The cisternae create a special environment for example in regard of calcium concentration which is critical for several cellular processes (Shibata et al., 2006). The ER was first described in the 1940s and one of the early electron microscopy ER images taken by the Nobel Prize winner George Palade is shown in Figure II (George E. Palade, 1966). It clearly shows the membrane organization of the ER and in addition the ER-associated ribosomes are visible. It was estimated that half of all ribosomes in a cell are in close proximity to the ER (Reid and Nicchitta, 2015). According to the dogma, these ribosomes should translate mRNAs encoding membrane and secretory proteins. However, it was shown in several studies that some mRNAs encoding cytosolic proteins are translated at the ER as well (Jagannathan et al., 2014; Mueckler and Pitot, 1982; Reid and Nicchitta, 2012; Voigt et al., 2017).

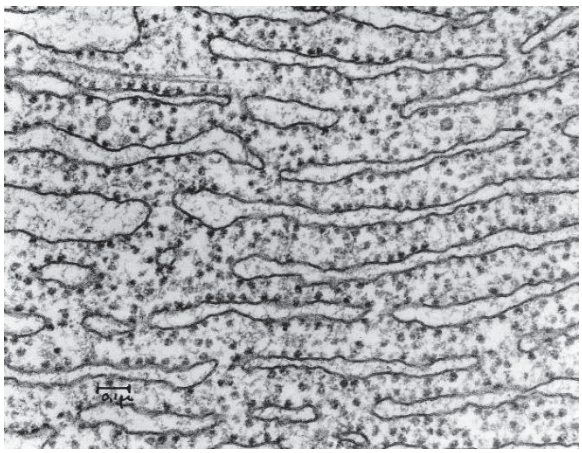

Figure II | Electron microscopy picture of the rough ER taken by Palade. The small black dots are ribosomes.

This raises the questions: How does a cell decide where an mRNA is translated and which factors influence that decision? How are ER-targeted mRNAs recognized and transported to the ER?

Targeting mRNAs to the ER

Major contributions to answer these questions came from the lab of Günter Blobel where the signal recognition particle (SRP) was discovered. The SRP is important for recognition and targeting of specific mRNAs co-translationally to the ER (see chapter "The signal recognition particle"). Blobel´s work was fundamental for our understanding of the secretory pathway and awarded with the Nobel Prize in 1999. Nevertheless, it has been shown that neither loss nor mutation of the SRP compromises the localization of numerous mRNAs to the ER (Costa et al., 2018; Lakkaraju et al., 2007; Mutka and Walter, 2001; Pyhtila et al., 2008) suggesting that multiple mechanisms for ER localization exist. Discovering these pathways is an outstanding challenge in the field and in recent years some progress has been made. In 2007, the TRC40/GET pathway was described explaining how tail-anchored proteins, which contain a single C-terminal transmembrane domain, are post-translationally targeted to the ER (Schuldiner et al., 2008; Stefanovic and Hegde, 2007). In addition, the SND (SRP-iNDependent targeting) proteins were identified in 2016 as an alternative targeting route to the ER

in yeast (Aviram et al., 2016). It was demonstrated that SND proteins are able to compensate loss of the SRP and it was suggested that they act as a backup system. The mammalian SND pathway is not fully characterized yet and not all human homologues have been identified (Aviram and Schuldiner, 2017). Overall, this shows that the fundamental question of how mRNAs are targeted to the ER is still not fully understood and that future research is needed to discover additional ER-targeting pathways.

Regulatory elements influencing ER-targeting

Besides the efforts to discover ER-targeting pathways, also the role of regulatory elements within mRNA sequences have been of high interest. Are there regulatory elements that contribute to ER-targeting? Blobel and colleagues proposed already in 1971 that proteins encode *cis* signals that determine their destination (Blobel and Sabatini, 1971) and it was shown that mRNAs encoding secretory proteins contain a N-terminal sequence encoding hydrophobic signal peptides that mediate SRP binding and ER-targeting. However, not all ER-targeted mRNAs contain these sequences but are still present at the ER.

Over the last decades, other regulatory elements were identified. It was observed that mRNAs encoding transmembrane proteins have a high uracil content suggesting that uracil-rich mRNAs are targeted to the ER. But further investigations showed that this feature is not sufficient to predict ER association (Cui and Palazzo, 2014; Prilusky and Bibi, 2009). In addition, regulatory elements influencing ER-targeting were described to be either located within the coding sequence (CDS) (Cui et al., 2013; Pechmann et al., 2014; Polyansky et al., 2013) or in the 3' untranslated regions (3'UTRs) (Chartron et al., 2016; Loya et al., 2008).

Recently, a global regulatory element occurring in the CDS but also in untranslated regions was identified (Cohen-Zontag et al., 2019). This *cis*-acting RNA sequence motif in ER-targeted mRNAs was termed SECReTE as secretion enhancing *cis*-regulatory targeting element and is present in humans and several eukaryotic organisms. The motif consists of up to 10 nucleotide triplet repeats enriched with pyrimidine (C and U) every third base. It is highly abundant in mRNAs encoding secretory proteins and it was shown in yeast that mutations within this motif impact ER localization, translation and secretion of these proteins. Furthermore, a screening in yeast was carried out to identify RNA-binding proteins that interact with SECeTE and several candidates were detected, but their role in protein secretion needs to be further determined (Cohen-Zontag et al., 2019).

Taken together, that suggest that *cis*-regulatory elements like SECReTE influence ER-targeting and downstream processing of membrane and secretory proteins. However, *trans*-acting factors that recognize such elements have not been characterized yet.

Translation at the ER

Targeting mRNAs to the ER is mainly mediated by the SRP (see next chapter) and coupled to translation. Therefore, in this section, the translation process will be introduced and the similarities and differences between cytosolic and ER translation will be highlighted. The translation of mRNA into protein is an extremely conserved, essential and very energy consuming process. It is highly regulated to ensure that the correct proteins are made in the proper places within the cell.

Similarities between cytosolic and ER translation

Translation can be subdivided in 4 main phases: initiation, elongation, termination and ribosome recycling (Schuller and Green, 2018) which are in principal identical between the cytosol and the ER. First, a complex consisting of eukaryotic initiation factors (eIF1, eIF1A, eIF3, eIF5), the initiator tRNA (Met-tRNA) and eIF2-GTP binds to the 40S small ribosomal subunit and forms the pre-initiation complex (PIC). PIC recognizes the cap of an mRNA that was preactivated by eIF4F and poly(A)-binding protein. After removal of secondary mRNA structures by RNA helicases PIC starts to scan the 5' untranslated region (5'UTR) for complementarity to the anticodon of the Met-tRNA. This scanning is promoted by eIF5. Once a start codon is found the 60S large ribosomal subunit joins PIC, which is catalyzed by eIF5B, and the 80S ribosome is complete (Hinnebusch, 2014). In total, it consists of 4 ribosomal RNAs (rRNAs) (18S, 28S, 5.8S, 5S) and 82 ribosomal proteins.

Next, elongation begins and GTP hydrolysis together with several eukaryotic elongation factors (mainly eEF1A, eEF2) enable the ribosome to move along the mRNA, 3 nucleotides per step. The mRNA itself consists of 3-letter codons that give rise to a total of 64 possible codons which encode 20 amino acids and stop signals. The codons are recognized and the appropriate transfer RNAs (tRNAs) delivers the corresponding amino acid which results in a growing peptide chain and later the final protein. The charged tRNA enters the A site (aminoacyl site) of the ribosome for GTP dependent peptide-bond formation (Figure III). Due to a conformational shift of the ribosome the tRNA moves to the P site (peptidyl site) and the amino acid is attached to the peptide chain. Lastly, the uncharged tRNA is moved to the E site (exit site) where it dissociates from the ribosome (Schuller and Green, 2018).

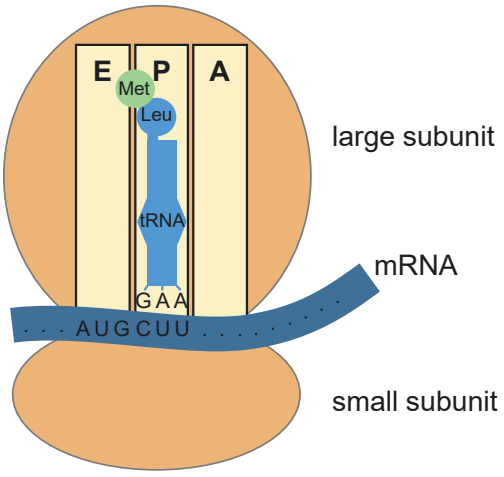

Figure III | Simplified scheme of the ribosome showing the E, P and A site.

This elongation process can be paused or stalled for multiple reasons like low levels of cognate charged tRNAs. The mRNA sequence itself can form secondary structures like pseudoknots resulting in ribosome pausing or the mRNA contains suboptimal codons leading to a supply limit of the corresponding tRNAs. Inhibitory conformations of the nascent peptide in the ribosome exit tunnel can also cause stalling. Then, either the machinery manages to continue translation or the ribosome becomes a target of quality control pathways (Buskirk and Green, 2017).
Once the ribosome reaches a stop signal the eukaryotic release factor eRF1 binds to the A site and with help of additional factors like eRF3 the nascent protein is released. Subsequently, the ribosome is recycled into the small and large ribosomal subunit primarily by ABCE1.

Differences between cytosolic and ER translation

Differences between cytosolic and ER translation remain incompletely explored. However, it was proposed that ER-associated translation is more efficient compared to cytosolic translation (Reid and Nicchitta, 2012). That is based on ribosome profiling of cell fractionated HEK293 cells that showed higher ribosome loading on ER-bound mRNAs compared to cytosolic mRNAs. In line with that, a recent single molecule imaging study detected more ribosomes on reporter mRNAs at the ER in comparison to the cytosol pointing out that cytosolic versus ER-associated translation are not identical and regulated differently (Voigt et al., 2017).
To understand how cytosolic and ER translation are regulated differently several aspects come into question.
First of all: do the ribosomes themselves differ in their composition? Ribosome heterogeneity was already suggested by Palade after he discovered the ribosomes (Palade, 1955) and is since then discussed in the field. Evidence is emerging that the ribosomes have distinct compositions in regard to the protein components but also regarding rRNA modifications (Genuth and Barna, 2018). A cryoelectron tomography study suggested that ER-associated polyribosomes have a specific three-dimensional arrangement which is different from cytosolic polyribosomes and which could explain why ER-associated ribosomes show higher translation efficiency (Pfeffer et al., 2012).
Secondly, specificity of ribosomes is also achieved by ribosome associated proteins. In a screen using mouse embryonic stem cells (mESCs) around 400 proteins were detected to be associated with the ribosome (Simsek et al., 2017). PKM2, a metabolic enzyme with RNA-binding ability, was discovered to interact with sub-pools of ER-associated ribosomes and to activate the translation of specific ER-targeted mRNAs. Nevertheless, no comprehensive differences regarding the composition, assembly or the translation states of ER-associated or cytosolic ribosomes have been identified.

Thirdly, RNA-binding proteins (RBPs) can contribute to translation regulation at the ER. LIN28A was shown to bind RNA and to inhibit translation of ER-targeted mRNAs in mESCs (Cho et al., 2012). Interestingly, LIN28A was also identified as a ribosome associated protein in the screen mentioned above (Simsek et al., 2017). Therefore, it is not clear if LIN28A achieves its specificity for ER-targeted mRNAs by binding to the mRNA and/or to the ribosome. Apart from LIN28A, no other RBPs have been described to regulate translation specifically at the ER.
In summary, only few aspects are already discovered to contribute to ER-specific translation and further ER translation regulators await characterization.

The signal recognition particle (SRP)

The signal recognition particle´s (SRP) major function is to facilitate co-translational ER-targeting of membrane and secretory proteins. It was discovered in the 1980´s and the core of the established SRP pathway remains unchanged to the present and will be summarized in this chapter. However, recent advances in technology have challenged some aspects of the SRP model and will be pointed out as well.

In eukaryotes, the SRP consists out of a ~300 nt long 7SL RNA scaffold and 6 proteins which are named according to their molecular weight (SRP9, SRP14, SRP19, SRP54, SRP68, SRP72) (Walter and Blobel, 1980). It is highly conserved and the bacteria SRP consists of a 4.5S RNA and only one SRP protein (SRP54/Ffh). This simplified SRP version is able to replace the eukaryotic SRP *in vitro* (Bernstein et al., 1993) demonstrating the evolutionary conservation of this pathway. According to the SRP pathway (Figure IV), the SRP recognizes and binds to its substrates in the cytosol during translation. The translation is stopped and the complex moves to the ER where the nascent protein chain is translocated into the ER lumen. The SRP dissociates and translation continues. In order to ensure proper function of this cycle there are several check points. Thereby, the specificity of the SRP is achieved which is crucial for the correct targeting of intrinsically insoluble SRP substrates.

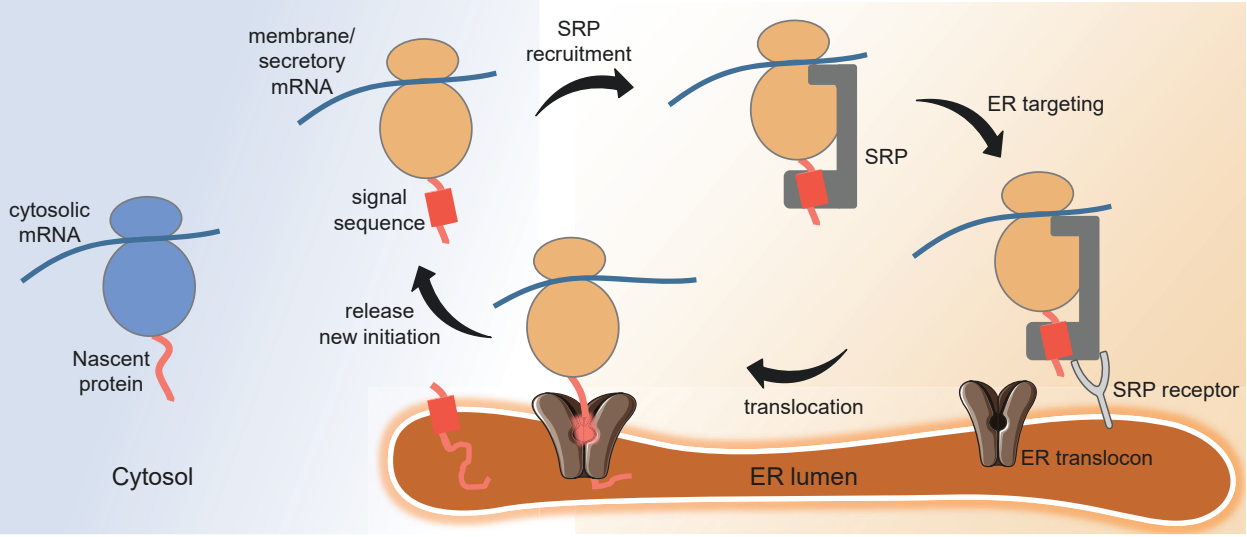

Figure IV | Schematic representation of the SRP pathway.

First, the SRP recognizes its substrates in the cytosol by a characteristic signal sequence that emerges from the ribosome exit tunnel. Secretory proteins contain a cleavable signal sequence, whereas membrane proteins have at least one non-cleavable transmembrane domain which has similar features as the signal sequence. The signal sequence is highly divergent regarding its length and amino acids composition but usually has a core of 8-12 hydrophobic amino acids, forms a α-helical structure and has basic residues at the N-terminus (Zhang and Shan, 2014). It needs to be recognized by the SRP before the translation exceeds a length of ~140 amino acids so that the SRP has a time window of around 3-5 seconds to recognize its substrate. Strong signal sequences recruit the SRP with a high affinity and mutation experiments showed that the binding affinity decreases by ~100 fold as the signal sequence is weakened (Zhang and Shan, 2014). Nevertheless, the SRP engages with strong and weak signal sequences at the ribosome exit tunnel and eliminates their exposure to the cytosol. Therefore, the hydrophobic groove of SRP54 accommodates the α-helical structure of the signal sequence and the methionine-rich M domain of SRP54 acts as a lid to additionally protect the signal peptides (Guna and Hegde, 2018).

Furthermore, SRP binding to the nascent chain complex leads to a translation arrest allowing the complex to move to the ER. The translation arrest was discovered by *in vitro* translation experiments using microsomal membranes representing the ER (Walter and Blobel, 1981). Fundamental for these experiments was the observation that *in vitro* translation of mRNAs containing a signal sequence results in a mature or immature protein depending on the presence of microsomal membranes. The mature protein is made when microsomal membranes are added to the reaction and is ~3 kDa smaller compared to the immature form. This size difference is due to the fact that the signal

sequence of the mature protein is cleaved by the signal peptidase which is part of microsomal membranes. In contrast, the ~3 kDa larger, immature protein still contains the signal sequence and is synthesized in absence of microsomal membranes. Making use of this observation an experiment was carried out where an mRNA was translated in the presence and absence of SRP and in addition microsomal membranes were added to the reaction at different time points. A simplified scheme of this experiment is shown in Figure V.

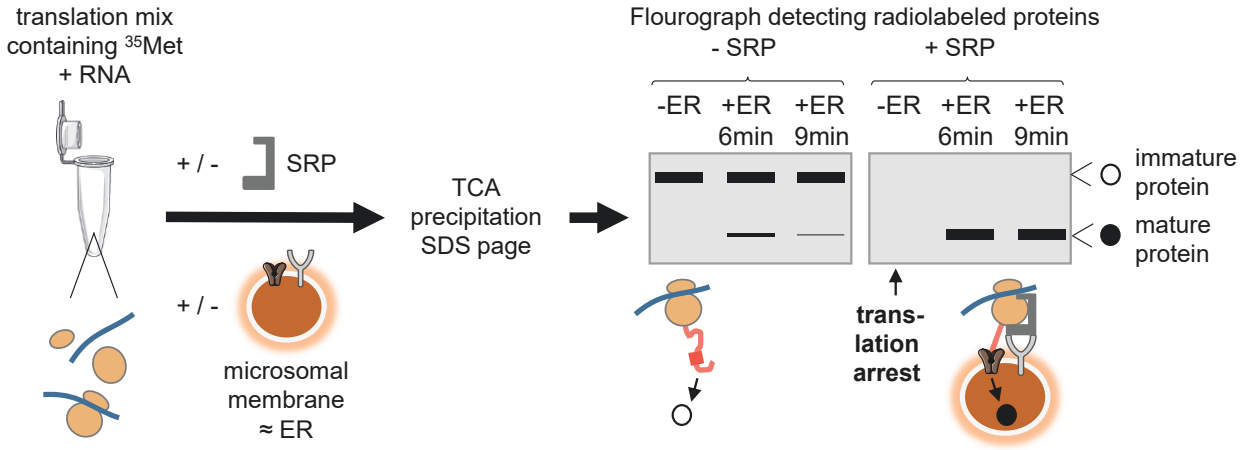

Figure V | Simplified scheme of *in vitro* translation experiments that proved that the SRP causes a translation arrest.

Strikingly, no protein product was observed in the reaction when the SRP was present but microsomal membranes were absent. Only upon addition of microsomal membranes the mRNA was translated and the mature protein was detected. In contrast, no translation arrest was observed in absence of SRP and microsomal membranes and only the immature protein product was detected. That proved that the SRP causes translation arrest which is undone upon addition of microsomal membranes. More than 20 years later the structure of the SRP bound to an elongation-arrested ribosome was solved and suggests that the Alu domain of the 7SL RNA is responsible for the translation arrest because it reaches the ribosomal binding site of elongation factors and thereby blocks their binding (Halic et al., 2004). This translation arrest allows the SRP complex to move to the ER and only correctly assembled complexes are steady enough to reach the ER. Incorrect complexes are up to 50 fold less stable (Zhang and Shan, 2014).

Once the SRP reaches the ER a GTP-dependent interaction with the SRP receptor takes place to unload the substrate. Unlike classical GTPases the SRP and SRP receptor are twin GTPases that do not need external GTPase activating proteins and instead activate each other by discrete conformational changes (Zhang and Shan, 2014). The coordinated progression of the conformational

changes is critical for the subsequent successful translocation of SRP substrates through the translocation channels e.g. Sec61 into the ER lumen. After GTP hydrolysis, the SRP dissociates back to the cytosol and the cycle starts over.

In summary, the specificity of the SRP is achieved through multiple checkpoints and SRP substrates are only correctly targeted to the ER if the SRP binds tightly to the signal sequence, the SRP complex is delivered quickly to the ER and GTP hydrolysis of the SRP - SRP receptor complex is well timed.

More recent findings

Recent technology advances and methods like ribosome profiling broadened our understanding of the SRP pathway. The client range of SRP was determined on a global scale by immunopurification of SRP and subsequent ribosome profiling to identify ribosome bound mRNA sequences in yeast (Chartron et al., 2016) and bacteria (Schibich et al., 2016). As expected, these experiments confirmed several points of the SRP hypothesis like high affinity of the SRP for hydrophobic amino acids within proteins harboring transmembrane domains. In addition, yeast ribosome profiling confirmed on the one hand that the SRP binds the ribosome-nascent-chain complex once the signal sequence is exposed. But on the other hand also ribosome-protected fragments upstream of the signal sequence were detected for many SRP substrates suggesting that the SRP can engage with the ribosome before the signal sequence is synthesized (Chartron et al., 2016). SRP pre-recruitment to the ribosome would have the advantage that the signal sequence can be bound faster after it emerges the ribosome exit tunnel. Moreover, SRP pre-recruitment provides an explanation how the SRP ensures punctual and correct binding to signal sequences. This is especially relevant considering the fact that translating ribosomes have a cellular concentration of 40-50 μM which is over 100 fold higher compared to the concentration of the SRP (Zhang and Shan, 2014). However, it is unknown how the SRP is pre-recruited to some but not other ribosomes and which factors mediate this binding.

These ribosome profiling experiments also started a debate about the relationship between SRP and translation. It was proposed that they did not confirm the initial observation from the 1980´s of a translation arrest once the SRP is bound to the ribosome-nascent-chain complex (Chartron et al., 2016; Guna and Hegde, 2018; Schibich et al., 2016). This statement is for example based on bacteria SRP selective ribosome profiling analysis. Here, the majority of SRP substrates were aligned according to their first SRP binding and the coverage after this binding was evaluated (Schibich et al., 2016). The authors argued that in theory this first SRP binding should mediate a translation arrest which should be detected by a peak in ribosome profiling data. However, this peak was not observed suggesting that the SRP is not causing a translation slowdown in bacteria. In line with that also the

yeast ribosome profiling results suggest that there is no classical translation arrest and that the elongation proceeds after the SRP binds (Chartron et al., 2016). This is based on cell fractionated ribosome profiling where first the membrane and soluble fraction were separated by centrifugation and subsequently ribosome profiling was carried out on both fractions. A schematic overview of this experiment and as well the results for one representative SRP substrate containing 4 transmembrane domains (OLE1) are shown in Figure VI.

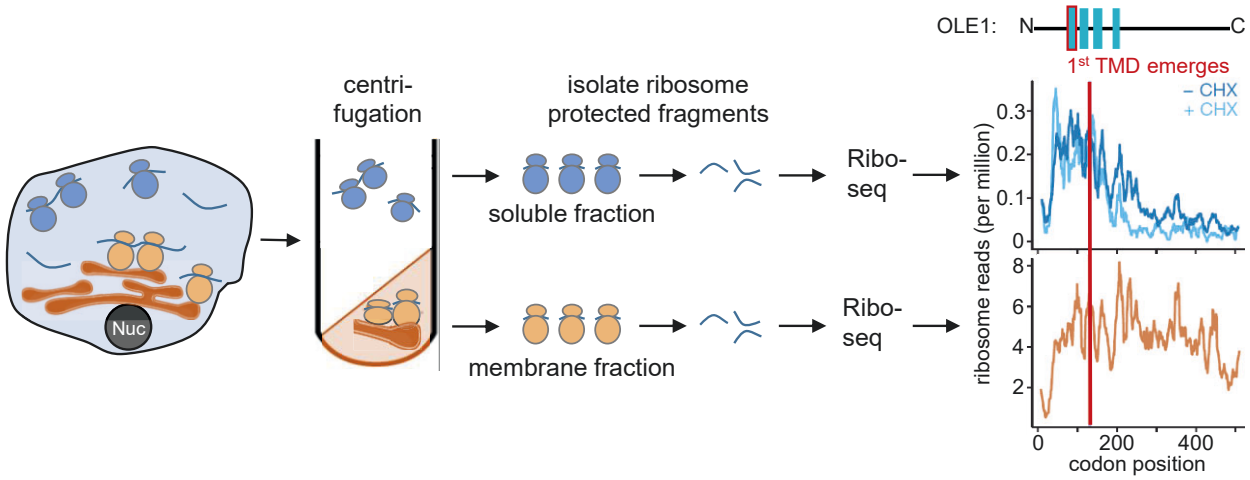

Figure VI | Simplified scheme of the experimental set up of yeast ribosome profiling and results for the transmembrane domain containing protein OLE1. Figure was reprinted with modifications from Chartron et al., 2016.

As expected the number of ribosome reads from the soluble fraction diminished after the targeting signal emerged from the ribosome exit tunnel. Nevertheless, some ribosome reads from the soluble fraction were still detected even after the targeting signal was fully translated suggesting that the SRP bound ribosome nascent chain complex remained soluble, elongation proceeded and therefore translation was not arrested. In summary, both yeast and bacteria ribosome profiling studies doubt the existence of an SRP-mediated translational arrest. However, this is based on the expectation that the first SRP substrate binding occurs directly after exposure of the signal sequence and immediately causes a translation arrest allowing no further translation in the cytosol. But clearly, the discovery that SRP pre-recruitment to ribosome nascent chain complexes exists, challenges this expectation. SRP pre-recruitment suggests that the SRP has alternative binding strategies besides the SRP hypothesis meaning that the SRP is not only binding its substrates when the first signal sequence is exposed but also earlier. Therefore, it can be speculated that SRP binding does not immediately lead to a translation arrest, but instead the translation arrest occurs at multiple positions along the mRNA. Thereby, it would be difficult to detect a translation arrest in ribosome profiling data because the ribosome profiling reads represent a global snapshot. Taken together, mechanistic details about the

translation arrest are not fully understood and it is not clear at which time point exactly the arrest occurs especially considering cases where the SRP is pre-recruited to its substrates. Likewise, it is unknown which other factors contribute to translational arrest.

Apart from challenging the SRP hypothesis regarding the translational arrest, the ribosome profiling experiments also opened up the question about de novo translation initiation at the ER. According to the current believes, the SRP only binds to free cytosolic ribosomes and at the end of an SRP cycle, the SRP and the ribosome dissociate back to the cytosol. This was already doubted and discussed in several articles (Hsu and Nicchitta, 2018; Jagannathan et al., 2014; Reid and Nicchitta, 2015) because it would exclude de novo translation initiation at the ER. The ribosome profiling experiment described in Figure VI now shows that ribosome-protected fragments are found in both the membrane and soluble fraction before the targeting signal is translated. According to the SRP hypothesis, no reads should be detected in the membrane fraction before the targeting signal emerged. However, these reads are observed and assuming that this is not due to cross contamination of the membrane and soluble fraction that would suggest that translation can be also initiated at the ER.

This conclusion is also supported by another study where ER selective ribosome profiling was carried out (Jan et al., 2014). To achieve ER specificity a biotin ligase was fused to an ER-resident protein and thereby only Avi-tagged ribosomes at the ER surface were biotinylated, purified and sequenced. Here, ribosome-protected reads were also observed before the targeting sequence was translated. Authors suggested a model wherein the pioneering round of translation facilitates ER-targeting of the ribosome-nascent-chain complex and afterwards the mRNA remains ER-associated and following ribosome are able to engage with the mRNA before the targeting signal is exposed (Jan et al., 2014).

Taken together, ribosome profiling revealed new insights into SRP function and opened up several points for clarification like further investigation of SRP-mediated translation arrest or mechanistic details about translation initiation at the ER.

The 15 KH domain-containing RBP HDLBP

The previous chapters highlighted that the secretory pathway is important for proper cell functioning and that, besides the well-established key points of the pathway, several mechanistic details are not fully understood. Certainly, there is evidence that additional ER-targeting pathways besides the SRP exist. Furthermore, *cis*-acting regulatory elements within the mRNA sequence have been identified to mediate ER-targeting, but no *trans*-acting factors like RNA-binding proteins are characterized to specifically bind these elements. In addition, mechanistic details within the SRP cycle and especially concerning the SRP-mediated translational arrest remain to be determined. Moreover, factors influencing ER-specific translation are unknown.

In order to contribute to uncover this lack of knowledge, we chose to study the RNA-binding protein HDLBP. Here, we will first introduce HDLBP and its characteristics and subsequently summarize the proposed functions of HDLBP and finally emphasize its disease relevance.

Characteristics

The protein HDLBP was first described in 1987 and it was found to bind to one of the major lipoproteins HDL (High-density lipoprotein) (Graham and Oram, 1987). Therefore, it was named high-density lipoprotein binding protein (HDLBP) which is the official gene name. HDLBP is also known as Vigilin due to the discovery of the chicken homolog in 1992 (Schmidt et al., 1992). The authors were not aware of the human homolog and therefore named the protein Vigilin based on the amino acids valine, isoleucine and glycine (VIG) reoccurring in the N-terminal part of the protein. HDLBP is conserved from yeast to human and ubiquitously expressed. Highest HDLBP mRNA expression in human tissues is found in secretory cells from e.g. the pancreas, the salivary gland or the liver (Figure VII).

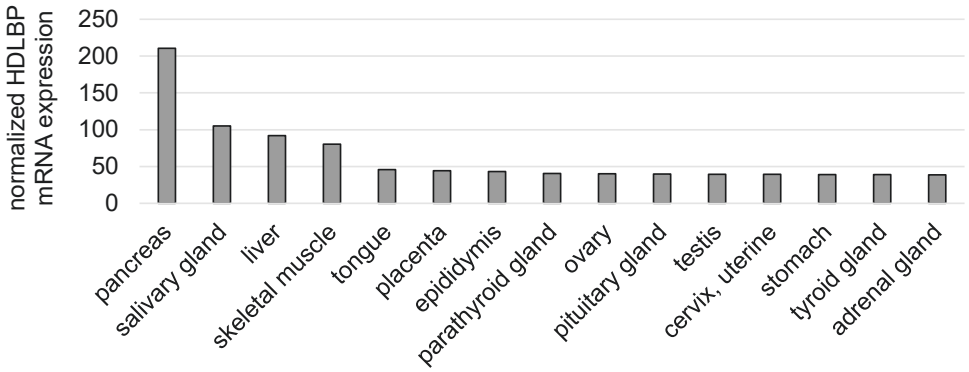

Figure VII | HDLBP mRNA expression is highest in secretory tissue. Consensus normalized expression levels for the top 15 human tissues with highest HDLBP mRNA expression are shown. Data are taken from proteinatlas.org (as of October 2020) where the transcriptomic datasets (GTEx, HPA, FANTOM5) were combined using a normalization pipeline.

Most characteristic for HDLBP are numerous hnRNP K homology (KH) domains. This domain was first discovered in hnRNP K, is highly conserved and comprises ~70 residues. The core motif of all KH domains is a minimal GxxG loop which is surrounded by alpha helices and ß sheets (Grishin, 2001). In eukaryotes, there are 2 additional alpha and ß elements at the C-terminus, whereas prokaryotes have additional elements N-terminal to this core motif (Nicastro et al., 2015). The GxxG loop typically allows binding of 4 unpaired bases of single-stranded DNA or RNA with micro molar affinity. Multiple copies of KH domains are believed to achieve more specificity of nucleic acid binding by for example producing larger binding sites (Valverde et al., 2008). Moreover, there are endogenous KH domains that have altered or interrupted GxxG core motifs which are classified as divergent KH domains and which most likely mediate protein interactions. Overall, 38 human RNA-binding proteins are annotated as KH domain containing protein and besides HDLBP all of them contain 1 to maximum 4 KH domains (RNA-binding protein database (rbpdb.ccbr.utoronto.ca) as of October 2020). In contrast, HDLBP has 15 KH domains making it an outstanding member of the KH domain family. Furthermore, HDLBP is also one of the largest RNA-binding proteins with a molecular mass of ~160 kDa. Human and mouse HDLBP contain 12 classical and 3 divergent KH domains (Figure VIII). Initially, HDLBP was reported to have 14 KH domains, but another divergent KH domain was found in the N-terminal part of the protein and numbered KH 0 (Cheng and Jansen, 2017). The structure of 7 human HDLBP KH domains is solved by crystallography (Protein database (rcsb.org) as of October 2020). The yeast homolog of HDLBP is called SCP160, has been intensively studied (see "Functions") and contains 14 KH domains of which 6 are classical and 8 are divergent (Figure VIII).

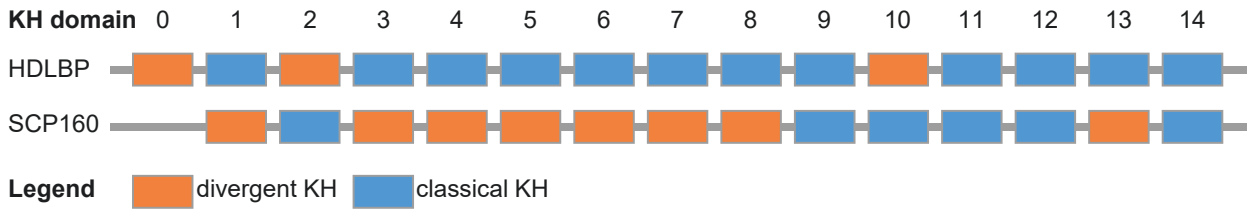

Figure VIII | Schematic overview of the KH domain composition of HDLBP and its yeast homolog SCP160. Figure was adapted from Cheng and Jansen, 2017.

Functions

Over the last decades of HDLBP research multiple functions of HDLBP and its homologs have been proposed. Even though there is no consensus about the exact function of HDLBP, there is no doubt that HDLBP binds nucleic acids. Most studies focus on the binding capacity of HDLBP to different types of RNA. Therefore, HDLBP is reported to influence many stages of RNA metabolism like tRNA and mRNA transport, translation or mRNA degradation. An overview of all the proposed functions of HDLBP can be found elsewhere (Cheng and Jansen, 2017). Here, we will summarize selected studies that illustrate what makes HDLBP a good candidate to be involved in the secretory pathway. These studies were mainly conducted in yeast and indicate on the hand that HDLBP contributes to ER mRNA targeting and on the other hand suggest that HDLBP is involved in translation regulation.

The role of HDLBP in mRNA transport to the ER

Only little direct evidence exists that HDLBP is involved mRNA targeting to the ER. However, it was shown for one selected Scp160p target that upon Scp160p deletion the mRNA signal was shifted from the membrane to the soluble fraction (Li et al., 2003). For another Scp160p target, it was observed that in pheromone-treated cells Scp160p was responsible for the correct localization of its target to the yeast-mating projection (shmoo) tip where it co-localizes with the cortical ER (Gelin-Licht et al., 2012). Furthermore, it was revealed that Scp160p binds to several mRNAs encoding polarity, secretion and mating factors and it was demonstrated that Scp160p mediates pheromone gradient sensing and mating response (Gelin-Licht et al., 2012). Nevertheless, it is unknown whether this is mechanistically due to targeting specific mRNAs to the cortical ER.

The role of HDLBP in translation regulation

Several studies in *S. cerevisiae* suggest that Scp160p associates with polyribosomes pointing to a function of HDLBP in translation regulation (Baum et al., 2004; Brykailo et al., 2007; Hirschmann et al., 2014; Lang and Fridovich-Keil, 2000; Lang et al., 2001; Li et al., 2004). Moreover, it was reported that Scp160p interaction with polyribosomes is mRNA dependent and that Scp160p -ribosome complexes are located at the ER (Frey et al., 2001). This is in agreement with the observation that Scp160p is associated with membrane-bound ribosomes (Weber et al., 1997) and suggests that Scp160p acts as an RNA-binding platform interacting with polyribosomes that are transported to the ER (Frey et al., 2001).

To uncover the role of Scp160p in translation, polysome profiling was performed upon Scp160p depletion and interestingly no global shift in translation was observed (Hirschmann et al., 2014).

Instead, Hirschmann et al. proposed that Scp160p controls the association of a specific subset of mRNAs with ribosomes and heavy polysomes. For one selected Scp160p target the study showed that Scp160p depletion led to translation downregulation and at the same time resulted in increased association with polysomes. It therefore seems that Scp160p is important for translation elongation (Hirschmann et al., 2014). In addition, a decrease for some ribosome-associated tRNAs was observed upon Scp160p depletion suggesting that Scp160p prevents the diffusion of discharged tRNAs or influences tRNA recharge (Hirschmann et al., 2014) which consequently affects translation.

Apart from these yeast studies, also a study using mice as a model organism showed that HDLBP influences translation. Moreover, this study revealed a role for HDLBP in liver metabolism (Mobin et al., 2016). Mobin et al. found in mouse hepatocytes that HDLBP binds to CU-rich regions in a subset of mRNAs encoding secreted proteins and enhances their translation. Direct evidence for this hypothesis was provided by *in vitro* translation assays whereby decreased translation of one HDLBP target (fetuin-A) was observed in HDLBP knockdown conditions. That resulted downstream in less production of feutin-A and subsequently in less secretion. Furthermore, less secretion of several other proteins was detected upon HDLBP knockdown and outstandingly also the secretion of VLDL (very-low-density lipoprotein) was decreased. VLDL is produced in the liver and mainly consists of apolipoproteins, cholesterol and triglycerides. It is transporting lipids throughout the body and a high blood level of VLDL is a known risk factor for cardiovascular diseases. To demonstrate the therapeutic potential of HDLBP Mobin et al. investigated the effect of hepatic HDLBP knockdown in in atherosclerosis prone Ldlr -/- mice. Importantly, they observed less atherosclerotic plaques upon hepatic HDLBP knockdown in Ldlr -/- mice showing that HDLBP is a promising target for cardiovascular research.

Taken together, several studies suggest that HDLBP is involved in translation regulation and especially Mobin et al. demonstrated that HDLBP influences the secretory pathway and showed that HDLBP enhances the translation of one mRNA target. However, the global effect of HDLBP on translation particularly in humans remains uncertain and to further exploit the therapeutic potential of HDLBP it is necessary to fully characterize its role in translation and the secretory pathway.

Disease relevance

As described above, HDLBP is a promising target for cardiovascular disease because it seems to influence the hepatic secretion of key lipoproteins like VLDL (Mobin et al., 2016). Furthermore, Mobin et al. reported that HDLBP is upregulated in livers of obese mice and in patients with fatty liver disease.

Moreover, it was shown that HDLBP is overexpressed in hepatocellular carcinoma and required for human hepatocellular carcinoma cell proliferation and tumor growth (Yang et al., 2014). Apart from hepatocellular carcinoma, HDLBP is also associated with lung adenocarcinoma (Kim et al., 2015), chronic myeloid leukemia (Cha et al., 2016) and breast cancer (Woo et al., 2011). Kim et al. sequenced 5 lung adenocarcinoma samples and found 2 samples with HDLBP mutations. Cha et al. analyzed by microarray chronic myeloid leukemia samples of 73 therapy responders and 38 non responders and found that HDLBP is differentially expressed between both groups suggesting that HDLBP might be useful as a diagnostic marker. Woo et al. investigated factors that modulate the expression of the proto-oncogene CSF1R in breast cancer and found that HDLBP inhibits CSF1R expression by binding to the 3'UTR. Subsequently, it was shown that overexpression of HDLBP in breast cancer cell lines led to less invasive behavior suggesting that HDLBP may function as a tumor repressor (Woo et al., 2011). Along these lines Molyneux et al. carried out a study to discover cancer genes by creating tumors with retroviral and transposon mutagenesis and found HDLBP as a candidate tumor repressor (Molyneux et al., 2014). Additional analysis revealed that there are 98 known mutations within HDLBP in patients with cancer and that HDLBP was commonly lost in several cancer types (Molyneux et al., 2014). Knockdown of HDLBP in human sarcomas cells resulted in higher induction of tumor development further supporting that HDLBP might act as a tumor repressor (Molyneux et al., 2014).

In addition, Felder and colleagues found out that HDLBP is down regulated in patients with autism and 2q37.3 deletion syndrome (Felder et al., 2009). Recently, Satterstom et al. identified in a large-scale exome sequencing study 102 genes that are associated with autism spectrum disorder and found among these genes HDLBP (Satterstrom et al., 2020).

Furthermore, HDLBP has been also associated with virus infection (Ooi et al., 2019). Ooi et al. screened for RNA-binding proteins that interact with the RNA of dengue and Zika viruses. Both viruses belong to the family of flavivirus and replicate in close association to the ER. HDLBP was discovered to interact with the viral RNA at the ER und it was suggested that HDLBP promotes optimal viral translation and replication (Ooi et al., 2019).

AIMS OF THE THESIS

The secretory pathway is essential for proper cell functioning and besides the well-established key points of the pathway, several mechanistic details are not completely understood. Especially the first steps of the pathway, where mRNAs encoding secretory and membrane proteins are targeted to and translated at the ER, remain to be further investigated. Fundamental questions like how are mRNAs transported to the ER and how is translation regulated at the ER are not fully answered yet.
Previous studies in yeast and mice suggested that the RNA-binding protein HDLBP contributes to these processes. The aim of this thesis is to characterize the human HDLBP and to understand its function.

Specific aims are:

- To identify RNAs bound to HDLBP
- To identify possible interactors of HDLBP
- To understand the role of HDLBP in transporting its targets to the ER
- To understand the role and function of HDLBP in translation

To study the function of HDLBP, we will use a variety of genomic and proteomic approaches applied to cultured human cells.

RESULTS

[1] Overview of generated cell lines

To study the function of HDLBP, we used as a cell culture model system the human embryonic kidney cell line HEK293 Flp-In T-Rex (HEK293). This cell line allows to easily generate stable cell lines with inducible expression of the protein of interest.

We made use of affinity tagging of HDLBP and generated HEK293 stable cell lines expressing either BirA-FLAG-HDLBP or FLAG/HA-HDLBP by hygromycin selection (Gregersen et al., 2014). The stable cell line stably expressing BirA-Flag-HDLBP was used for BioID experiments (see chapter 6). For PAR-CLIP, immunoprecipitation and polysome profiling experiments, we took advantage of the cell line stably expressing FLAG/HA-HDLBP (see chapter 2, 6, 8).

For most other experiments we used HEK293 HDLBP knockout (HDLBP KO) cell lines in comparison to wild-type (WT) HEK293 cells. HDLBP KO cell lines were generated with the CRISPR-Cas9 system using two different guide RNAs. Guide RNA 1 (gRNA1) targeted exon 11 and guide RNA 2 (gRNA2) targeted exon 4.

To understand the impact of HDLBP on cancer biology, we used as a cell culture model system the lung carcinoma cells A549 (see chapter 11). We also generated A549 HDLBP knockout cells, but only guide RNA 2 resulted in surviving colonies.

Overall, A549 and HEK293 HDLBP KO cells were viable showing that HDLBP is a non-essential protein in both cell types which is consistent with findings from human chronic myelogenous leukemia cells (KBM7) (Blomen et al., 2015; Wang et al., 2015) and human hepatoma cells (Huh 7.5.1) (Ooi et al., 2019).

Western Blot analysis of all cell lines used in this study is shown in Figure 1.

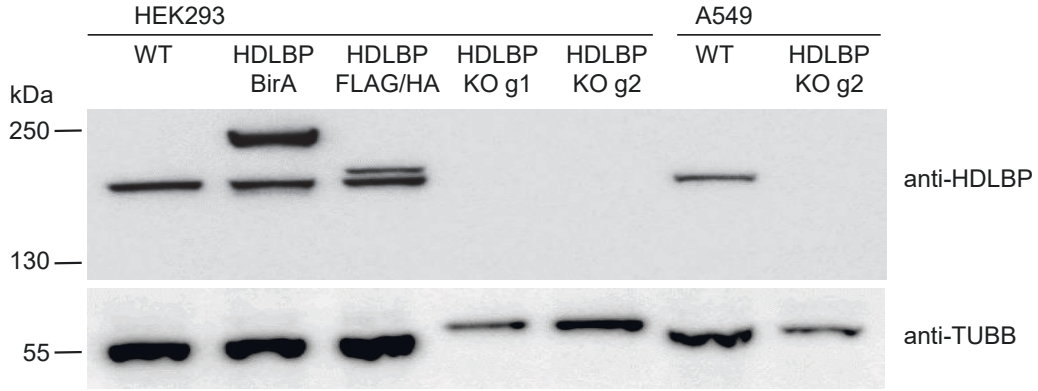

Figure 1 | All HEK293 and A549 cell lines used in this study are shown. Western Blot analysis was performed with a primary antibody against HDLBP and an anti-tubulin antibody was used as a loading control.

[2] HDLBP is a cytoplasmic protein in HEK293 cells that co-localizes with the ER

To determine function of HDLBP, we first investigated the subcellular distribution of HDLBP in our human cell system, in HEK293 cells. Previous findings suggest that HDLBP is a cytoplasmic protein (Neu-Yilik et al., 1993) and the human protein atlas lists the main localization of HDLBP as cytosol which is supported by immunohistochemistry data from A-431 cells (epidermoid carcinoma cells) , U-2 OS cells (osteosarcoma cells) and U-251 MG cells (glioblastoma cells) (proteinatlas.org as of October 2020). It was shown that HDLBP also co-localizes with the ER in Huh 7.5.1 cells (hepatoma cells) (Ooi et al., 2019) and in HEK293 cells (Lu et al., 2012). Besides that, a small fraction of HDLBP was also reported to be present in the nucleus (Kügler et al., 1996; Vollbrandt et al., 2004; Wang et al., 2005; Zhou et al., 2008).

To clarify the localization of HDLBP in our cell system, we performed immunohistochemistry experiments using an antibody against HDLBP in combination with an ER-tracker and Hoechst nucleus staining. We show that HDLBP is localized in cytoplasm of HEK293 cells and also co-localizes with the ER (Figure 2).

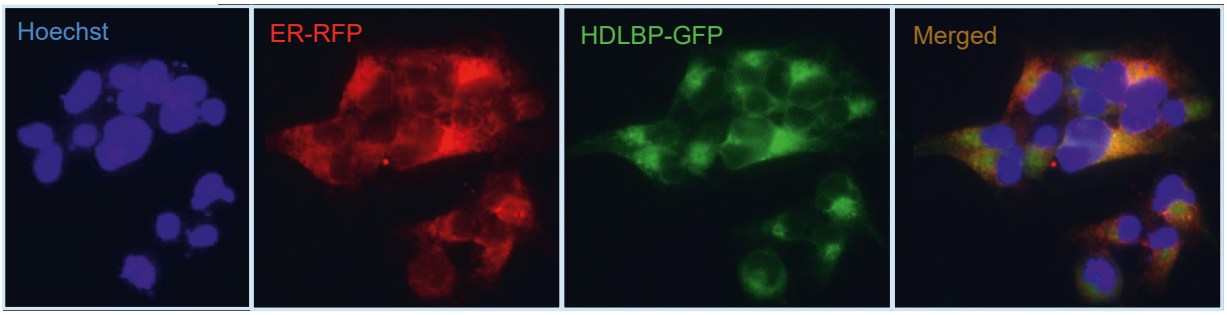

Figure 2 | Representative immunohistochemistry pictures of HEK293 cells stained with HDLBP (green), ER-tracker (red) and Hoechst (blue) showing that HDLBP is a cytoplasmic protein that co-localizes with the ER.

In addition, we carried out cell fractionation studies where we biochemically separated the cytoplasmic and membrane fraction (see chapter 4 for details). To confirm that this separation worked, we analyzed the localization of marker proteins with known localization (ER: BCAP31, cytosol: Tubulin) and found that the marker proteins distributed as expected (Figure 3). HDLBP was present in both fractions and enriched in the membrane fraction.

We conclude that HDLBP is a cytoplasmic protein in HEK293 cells that co-localizes with the ER.

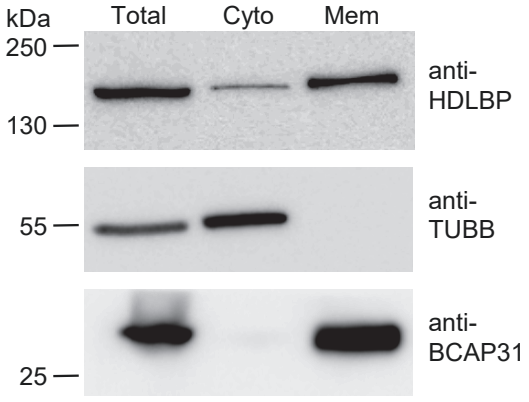

Figure 3 | Western Blot analysis of cell fractionation purity using marker proteins (ER: BCAP31, cytosol: TUBB). HDLBP is enriched in the membrane compartment.

[3] HDLBP-bound mRNAs encode membrane and secretory proteins

In the interest of understanding the function of the RNA-binding protein HDLBP it is crucial to identify RNA targets of HDLBP. Previously, it was shown that the mouse homolog of HDLBP mainly binds mRNAs encoding secretory proteins in hepatocytes which may have therapeutic potential for cardiovascular diseases (Mobin et al., 2016). To further exploit this therapeutic potential, we characterized the RNA targets of the human HDLBP by PAR-CLIP and validated our findings by immunoprecipitation experiments in combination with real-time PCR.

HDLBP PAR-CLIP

To identify direct RNA targets of HDLBP, we performed PAR-CLIP (Hafner et al., 2010) using a HEK293 cell line that stably expressed FLAG/HA tagged HDLBP (see chapter 1, Figure 1). We observed a similar expression level of FLAG/HA tagged HDLBP compared to endogenous HDLBP. Nascent RNA of these cells was metabolically labelled with 4-thiouridine (4sU) and by UV crosslinking the proteins were covalently linked to the labelled RNA. HDLBP-RNA complexes were immunopurified, radiolabeled and gel separated (Figure 4). HDLBP-bound RNA fragments were extracted, ligated to linkers and converted to a cDNA library. The 4sU incorporation caused T-to-C conversions during reverse transcription which allowed us to specifically identify HDLBP crosslinking sites on the target RNAs. Illumina sequencing of 2 PAR-CLIP replicates was carried out and after processing the reads, we only kept reproducible read clusters.

We observed as expected a high frequency of T-to-C conversions. To obtain a HDLBP PAR-CLIP enrichment value per gene, we summed up all reproducible T-to-C conversions per gene and corrected this value for expression level bias. In addition, we assessed the distribution of T-to-C conversions within the 5’ untranslated regions (5’UTR), the coding sequence (CDS) and the 3' untranslated regions (3'UTRs) of mRNAs.

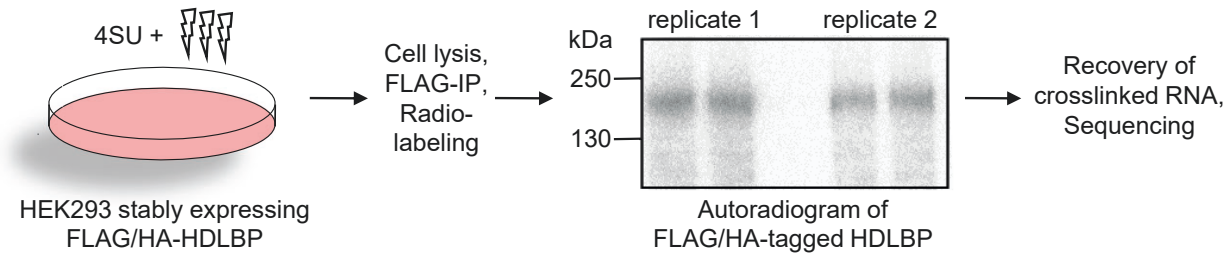

Figure 4 | Experimental outline of the PAR-CLIP protocol with an autoradiograph of ^{32}P-labeled RNAs crosslinked to HDLBP.

We found that the majority of T-to-C conversions (~50%) was present in the CDS of mRNAs (Figure 5A). Around 20% of crosslinking signal was detected in 3'UTRs. We also identified crosslinks to the 5'UTR, ribosomal RNA (rRNA), transfer RNA (tRNA) and SRP-RNA (7SL RNA). A minority of reads was detected in introns and intergenic regions.

To characterize HDLBP-bound mRNAs, we compared HDLBP PAR-CLIP binding enrichment for different mRNA classes based on the presence of known targeting signals in corresponding proteins (Figure 5B). For cytosolic, mitochondrial or tail anchored transmembrane mRNAs we did not detect HDLBP binding enrichment. Modest HDLBP binding enrichment was observed for membrane-bound mRNAs with no known targeting signal (TS). In contrast, we found highest HDLBP binding enrichment for mRNAs encoding a signal peptide (SignalP) or transmembrane domains (TM) or both.

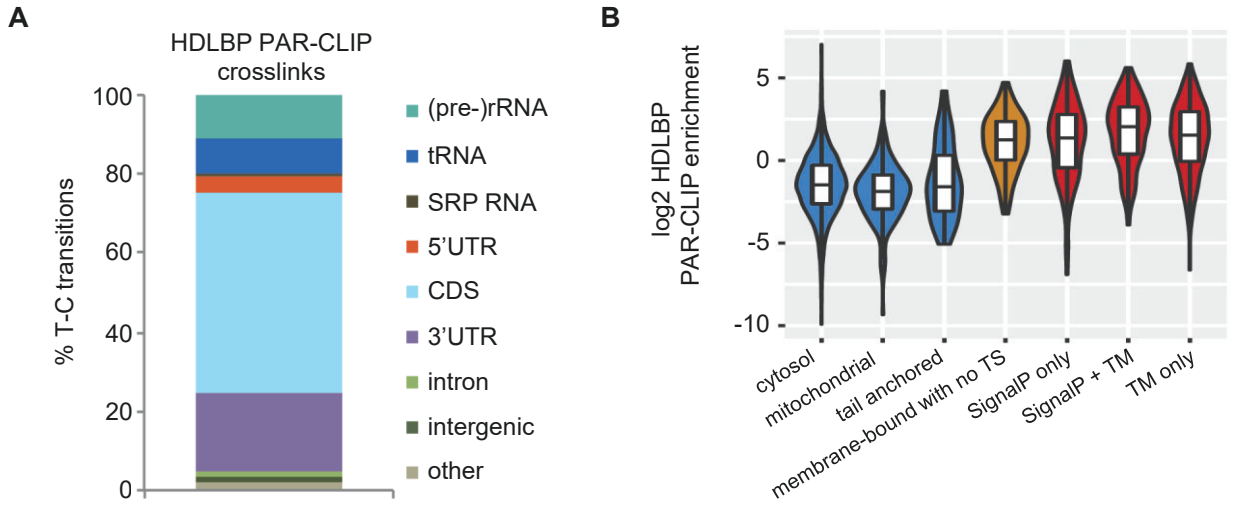

Figure 5 | HDLBP mainly binds the CDS of mRNAs that encode a signal peptide (SignalP) or a transmembrane domain (TM) or both. **(A)** Distribution of HDLBP binding sites (T-to-C transitions) in different RNA-species. **(B)** HDLBP PAR-CLIP binding enrichment for several mRNA classes.

PAR-CLIP validation

To validate the PAR-CLIP findings, we carried out RNA immunoprecipitation (RIP) experiments. Therefore, we used as well HEK293 cells stably expressing FLAG/HA tagged HDLBP (see chapter 1, Figure 1). For immunoprecipitation beads were coupled to either FLAG antibody or to an IgG control. FLAG/HA tagged HDLBP and its associated RNAs were enriched and the RNA was isolated. The RNA was reverse transcribed and detected by real-time PCR. Several HDLBP mRNA targets showed as expected high enrichment, whereas our HDLBP non-target control, a mitochondrial DNA encoded mRNA (MT-CO1), showed no enrichment (Figure 6). Characteristics of the validated HDLBP mRNA targets and their cellular mRNA localization (see chapter 4) are shown in Table 1. In addition, we found 7SL RNA, which is part of the SRP complex, moderately enriched. This validation showed that

the crosslinking signal detected by PAR-CLIP results from stable interactions of HDLBP with its RNA targets.

Gene	*mRNA localization (our data)*	*Feature*
ATP1A1	membrane	10 transmembrane domains
ATP1A3	membrane	10 transmembrane domains
CD46	membrane	1 transmembrane domain
CD71 / TFRC	membrane	1 transmembrane domain
HNRNPUL1	cytosol	-
IGF2R	membrane	1 transmembrane domain
LAMB1	membrane	signal peptide
LAMC1	membrane	signal peptide
PFN1	cytosol	-
TMBIM6	membrane	6 transmembrane domains
YWHAZ	cytosol	-

Table 1 | HDLBP mRNA targets identified by PAR-CLIP and validated by immunoprecipitation coupled with real-time PCR.

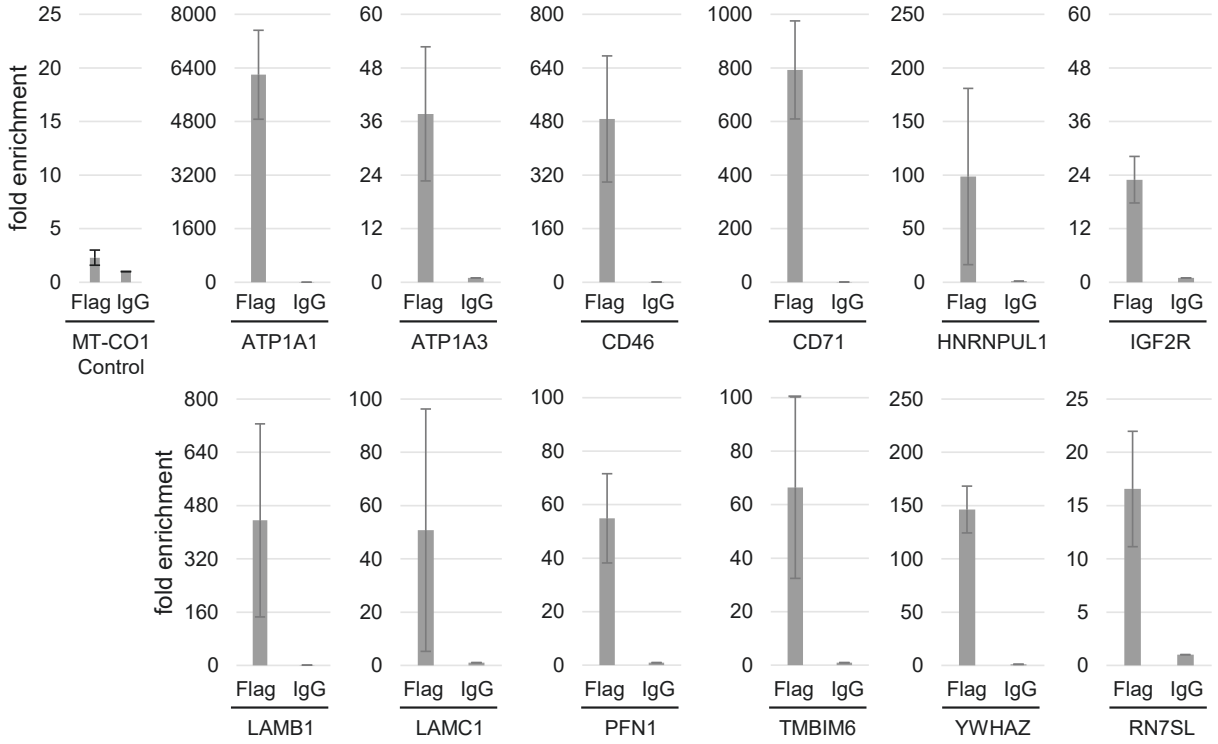

Figure 6 | Fold enrichment (anti-FLAG vs. IgG control) for several HDLBP targets and a negative control (MT-CO1) as detected by real-time PCR. The average of 4 replicates is shown and the corresponding standard deviation as indicated by the error bar.

Taken together, we obtained a deep HDLBP PAR-CLIP data set which we validated by immunoprecipitation coupled with real-time PCR. We found that HDLBP binds several RNA species but mainly to the CDS of mRNAs that encode for secretory and transmembrane containing proteins.

[4] HDLBP interacts with ~80% of all ER-localized mRNAs

HDLBP mRNA targets mainly encode for secretory and TM containing proteins and are known to be localized and translated at the ER. We next wanted to understand to which extent HDLBP is binding ER-localized mRNAs. Previously, it was reported that the mouse homolog of HDLBP is only binding a specific subset (17,5%) of mRNAs encoding secretory proteins in hepatocytes (Mobin et al., 2016). However, the specificity of HDLBP in human and non-secretory tissue is unknown.

Here, we performed cell fractionation experiments and mRNA sequencing to first classify ER- and cytosol-localized mRNAs. The experimental outline is shown in Figure 7. We biochemically separated the cytoplasmic and membrane fraction by sequential detergent extraction (Jagannathan et al., 2011). Cells were first permeabilized with a digitonin containing buffer. This allowed to specifically solubilize the plasma membrane because digitonin is a ß-sterol binding detergent and the plasma membrane has a higher cholesterol content compared to other membranes. The ER and nuclear membranes stayed intact and only the cytosolic components were collected. Next, a NP-40 containing buffer was used to further solubilize the ER and mitochondria membranes but leaving the nuclei, the intermediate filament network and the actin cytoskeleton intact which allowed us to collect the membrane fraction.

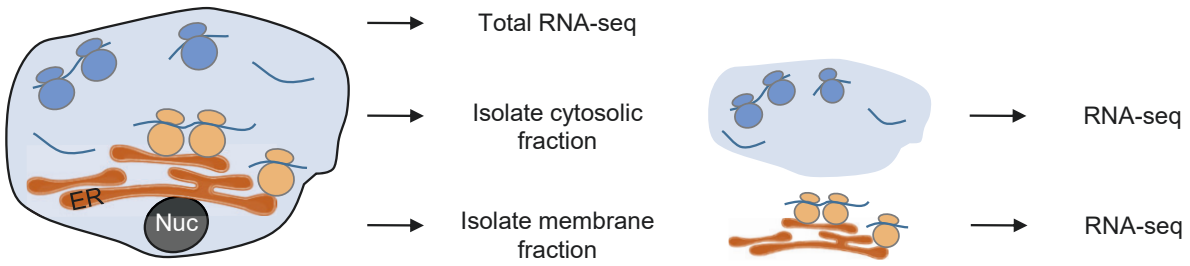

Figure 7 | Overview of the cell fractionation experiment in combination with RNA-sequencing.

We confirmed the successful separation of cytosol and membrane fractions by Western Blot analysis of proteins with known localization (see chapter 2, Figure 3, ER marker: BCAP31, cytosol marker: Tubulin). RNA was isolated from both fractions and in addition from non-fractionated (total) cells. RNA-sequencing libraries were prepared and sequenced. A principal component analysis was carried and showed that membrane, cytosol and total fractions are separated from each other as expected (Figure 8A). Replicate samples were highly reproducible as shown in the correlation matrix confirming the high quality of the data set (Figure 8B).

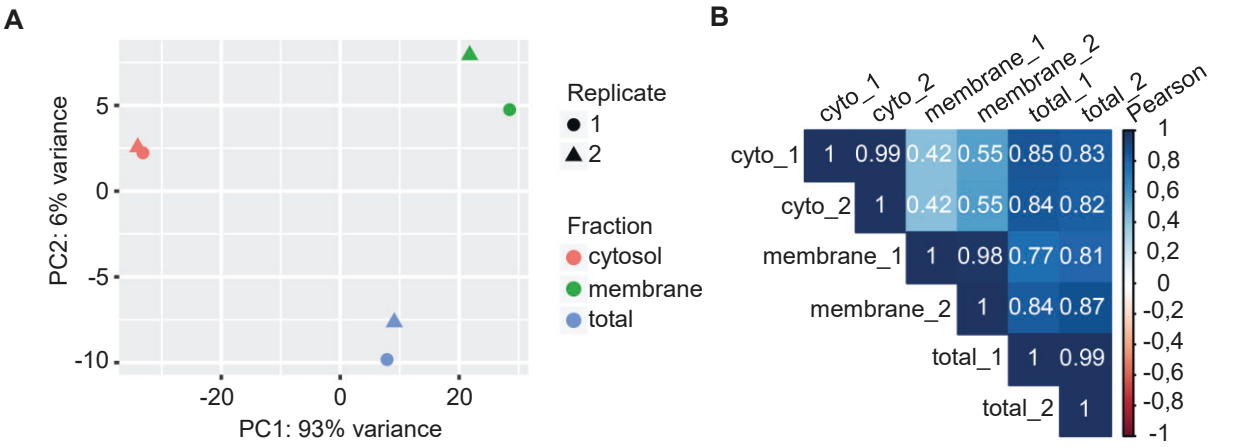

Figure 8 | Quality control of the cell fractionation experiment in combination with RNA-sequencing. **(A)** Principal component analysis using variance stabilized transformation of RNA-sequencing read counts per gene. **(B)** Correlation matrix of pairwise comparisons between non-transformed RNA-sequencing read counts per gene derived from cytosol, membrane or total fractions. The Pearson coefficients are shown.

We found mRNAs enriched in the cytosol and in the membrane fractions which enabled us to classify mRNAs according to their localization (Figure 9A). In total we classified 7292 mRNAs: 6135 cytosol-localized mRNAs and 1157 membrane-localized mRNAs. 193 mRNAs did not show specific enrichment for any compartment. As described above the membrane fraction includes mRNAs localized at the ER membrane but in addition also mitochondrial mRNAs. Therefore, we filtered our 1157 membrane-localized mRNAs for mitochondrial encoded mRNAs. We found that all 13 protein coding mitochondrial mRNAs (Barchiesi and Vascotto, 2019) were, as expected, highly enriched in our membrane fraction. Overall, our membrane fraction mainly consists of ER-localized mRNAs (despite the 13 mitochondrial mRNAs) and in consequence we also use the term ER-localized mRNAs. To further explore our defined groups, we analyzed them for the enrichment of mRNAs that encoded or lacked co-translational targeting signals. We found highly enriched in the ER-localized mRNA class as expected mitochondrial DNA encoded proteins, mRNAs encoding a SignalP or TM and mRNAs encoding both (SignalP and TM) (Figure 9B). In contrast, mRNAs encoding post-translationally targeted tail-anchored transmembrane proteins and mRNAs encoding nuclear DNA encoded mitochondrial proteins were not enriched in the membrane fraction as expected giving us high confidence about our classification.

To determine the localization of HDLBP mRNA targets, we plotted HDLBP crosslinking signal (T-to-C conversions) versus the mRNA abundance in order to normalize for the expression level. We colored each mRNA according to our classification (orange = ER-localized, blue = cytosol-localized) (Figure 9C). Strikingly, strong HDLBP mRNA targets were mainly ER-localized. We estimated that more than 80% of all ER-localized mRNAs were bound by HDLBP (Figure 9D).

In summary, we obtained a high quality mRNA sequencing data set of fractionated HEK293 cells which allowed us to estimate that HDLBP binds more than 80% of all ER-localized mRNAs.

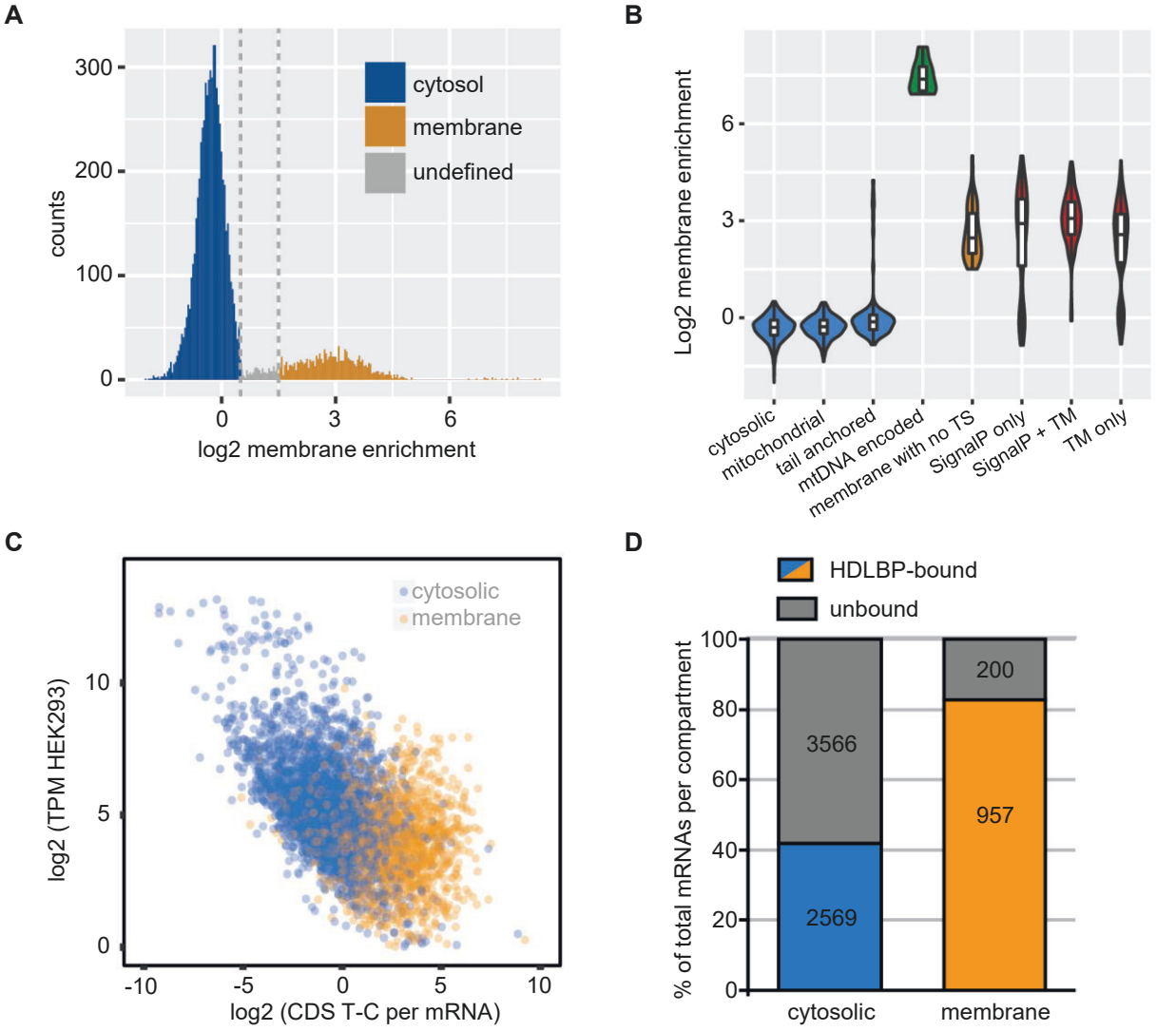

Figure 9 | HDLBP binds to ~80% of all ER-localized mRNAs.
(A) Quantification of steady state membrane and cytosolic localization of mRNAs by cell fractionation experiments in combination with RNA-sequencing. The histogram shows the log2-transformed membrane enrichment which was defined as the ratio between membrane versus cytosolic read counts only including mRNA with a TPM of at least 10. Membrane bound/ER-localized mRNAs (n=1557) and cytosolic mRNAs (n=6135) were classified using cutoffs of 0.5 and 1.5 (grey lines).
(B) Quality control of RNA-sequencing derived classification of mRNA localization. Log2-transformed membrane enrichment is shown for different mRNA classes which were defined according to published data (see Materials and Methods "Targeting signal annotations").
(C) Scatter plot of expression-normalized HDLBP PAR-CLIP CDS crosslinking signal versus mRNA expression level in non-fractionated/total cells. Each dot represents an mRNA and the color of the dots correspond to the mRNA localization as classified in (A). Membrane-localized mRNAs are shown in orange and cytosol-localized mRNAs are colored in blue.
(D) HDLBP binding to membrane and cytosolic mRNAs as classified in (A). Percentages and absolute numbers of HDLBP mRNA targets within membrane or cytosol-localized mRNAs are given.

[5] HDLBP binds to a CU-containing motif in the CDS of ER-localized mRNAs and forms multivalent interactions

After defining mRNA targets of HDLBP, we next wanted to understand the binding preferences of HDLBP within mRNA transcripts. Previously, it was reported that RNA recognition elements of the mouse homolog of HDLBP are a tandem of CHHC or CHYC (H = A/C/U and Y = C/U) spaced by 2-8 nt occurring evenly distributed along the CDS as determined by PAR-CLIP of primary hepatocytes (Mobin et al., 2016). *In vitro* assays using the xenopus homolog of HDLBP and the human HDLBP identified the binding motif as (A)nCU and CU(A)n within a 75 nt long single stranded region (Kanamori et al., 1998). However, the detailed binding characteristics of the human HDLBP are unknown. Here, we show the binding preferences of HDLBP within mRNA transcripts, we uncover the binding motif of HDLBP, investigate HDLBP binding to longer RNA recognition elements and characterize multivalent HDLBP interactions.

Binding preferences of HDLBP within mRNAs

To assess the binding preferences of HDLBP from our PAR-CLIP data, we first normalized the reproducible T-to-C conversions within cytosol- and ER-localized mRNA targets for the library size. Next, we grouped the HDLBP crosslinks according to their occurrence in 5'UTR, CDS and 3'UTR. This grouping was highly reproducible between replicates (Figure 10) ensuring us that we can extract binding preferences of HDLBP from the PAR-CLIP data.

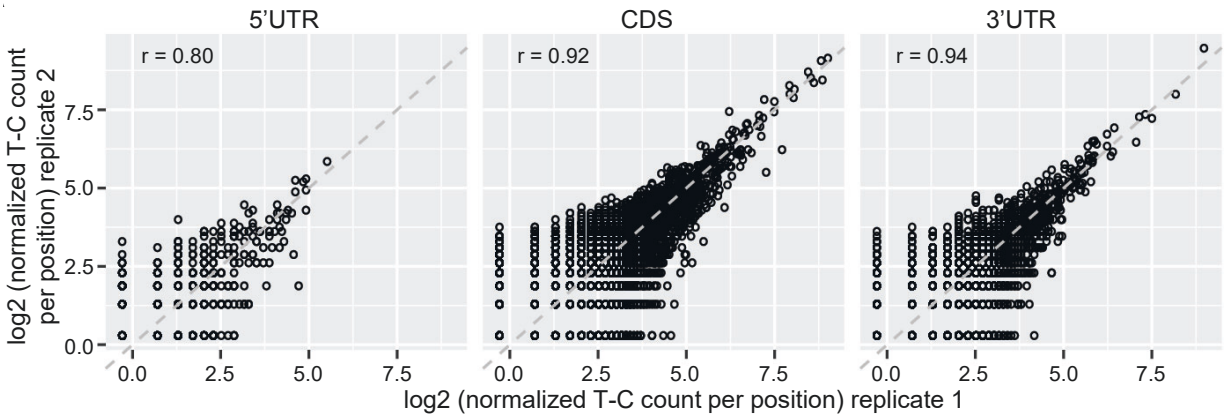

Figure 10 | PAR-CLIP derived T-to-C conversions within mRNA transcript regions are highly reproducible between replicates. Scatter plots show normalized T-to-C conversions per million for 2 HDLBP PAR-CLIP replicates as detected in the 5'UTR, CDS and 3'UTR. The Pearson correlation coefficients are shown.

Subsequently, we generated a high resolution map of HDLBP crosslinks within transcript regions of cytosol- and ER-localized mRNAs (Figure 11). Therefore, we included transcripts with at least 5 T-

to-C conversions per million reads and plotted at each position the mean scaled T-to-C signal. We observed more 5'UTR HDLBP binding in cytosol-localized mRNAs but overall we only detected less than 5% of HDLBP crosslinks in the 5'UTR (chapter 3, Figure 5A). Interestingly, we found that HDLBP predominantly binds all over the CDS of ER-localized mRNAs. In contrast, HDLBP binding to the 3'UTR was more prevalent in cytosol-localized mRNAs.

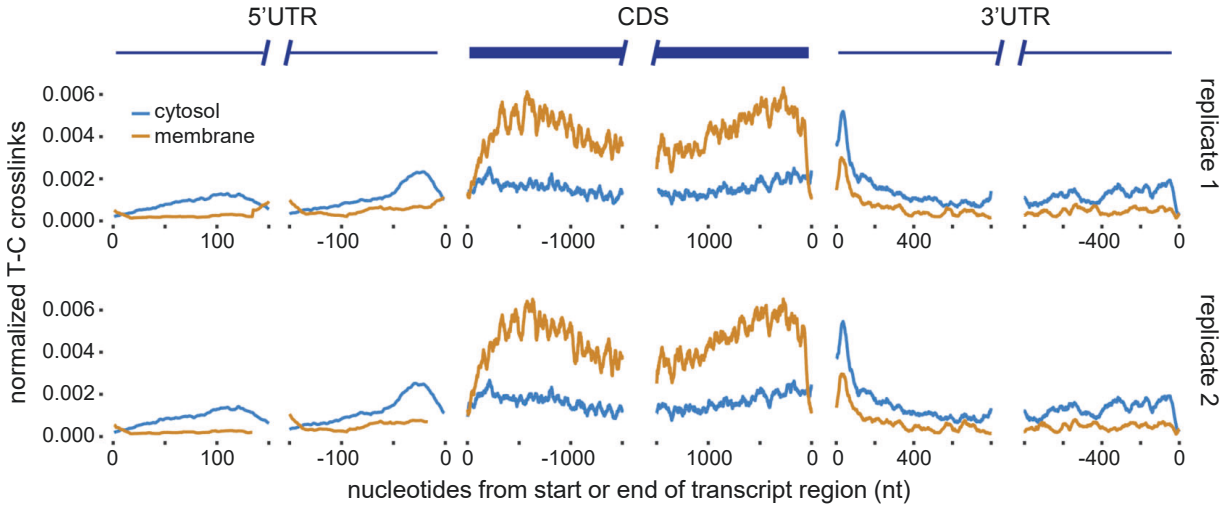

Figure 11 | HDLBP binds ER-localized mRNAs all over the CDS. Meta-transcript analysis of HDLBP crosslinking signal for cytosol-localized mRNAs (blue) and ER-localized mRNAs (orange) as classified by cell fractionation experiments in combination with RNA-sequencing (chapter 4).

To further characterize HDLBP binding to the CDS versus 3'UTR, we analyzed the ratio between length-normalized T-to-C conversions detected in the CDS versus 3'UTR in ER- versus cytosol-localized mRNAs. ER-localized HDLBP mRNA targets showed higher CDS/3'UTR ratios compared to cytosol-localized mRNA targets (Figure 12A).

To understand where we detected most of the HDLBP crosslinking signal, we defined similar sized groups of ER- and cytosol-localized mRNA targets. Groups were composed of mRNAs with high (T-to-C conversions > 1.39), mid (T-to-C conversions 0.3 – 1.39) and low (T-to-C conversions < 0.3) HDLBP crosslinks. The CDS versus 3'UTR ratio of each group was plotted and showed that most HDLBP crosslinks are detected in the CDS of ER-localized mRNA (Figure 12B).

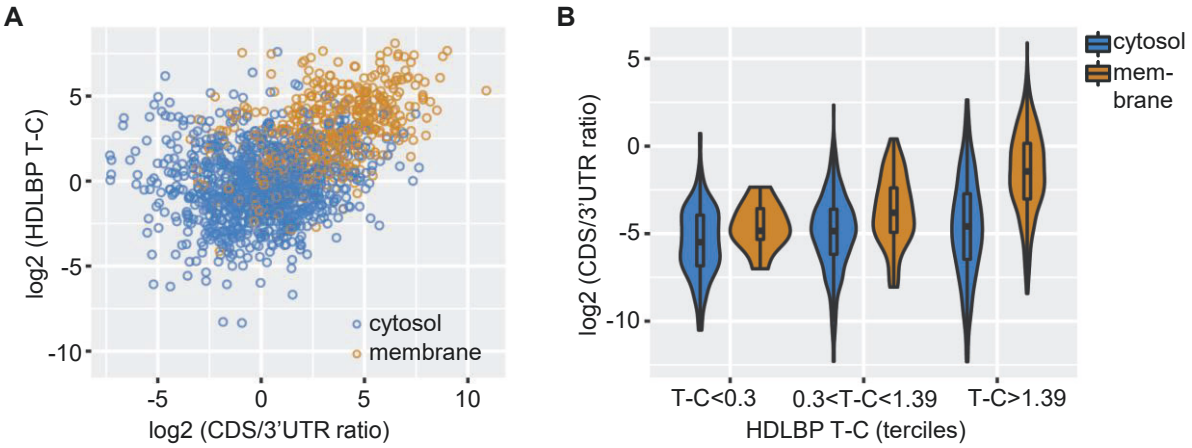

Figure 12 | HDLBP mainly binds the CDS of ER-localized mRNAs. **(A)** Scatter plot of the CDS/3'UTR ratio of HDLBP crosslinks plotted against total HDLBP crosslinks (membrane mRNAs = orange, cytosol mRNAs = blue). **(B)** Groups with high, mid and low HDLBP crosslinks were plotted against the distribution of the CDS/3'UTR ratio.

In order to support this global analysis, we selected an ER-localized (IGF2R) and a cytosol-localized (HNRNPUL1) HDLBP mRNA target (also validated by real-time PCR, see Figure 6) and visualized the PAR-CLIP read coverage and T-to-C conversions (Figure 13). For the cytosol-localized mRNA HNRNPUL1, we detected on average equal read coverage in the CDS and 3'UTR. The ER-localized HDLBP target IGF2R showed high read coverage in the CDS confirming our conclusion that HDLBP predominantly binds the CDS of its ER-localized mRNA targets.

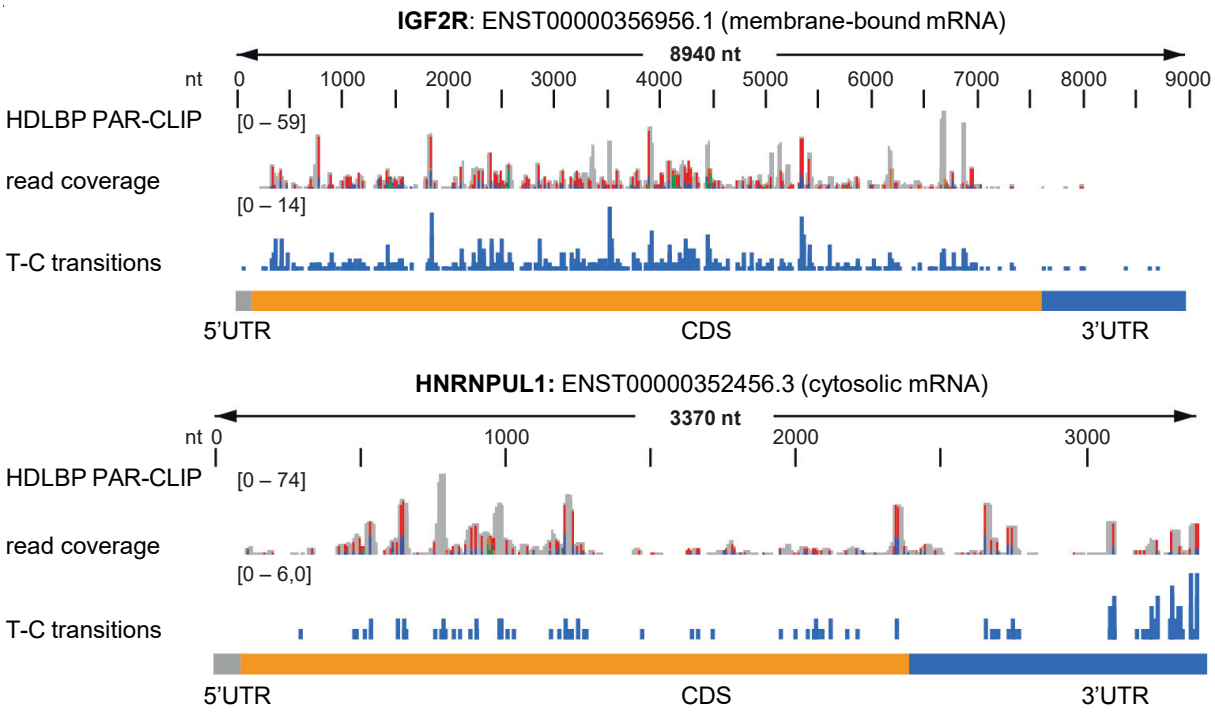

Figure 13 | Browser representation of HDLBP PAR-CLIP read coverage and T-to-C conversions for IGF2R mRNA (membrane-bound HDLBP target) and HNRNPUL1 mRNA (cytosolic HDLBP target).

Binding motif of HDLBP

Next, we wanted to identify primary sequence characteristics of HDLBP binding. Therefore, we generated for each crosslinking position 7-mers. These 7-mers were ranked by their frequency relative to all crosslinked 7-mers (Figure 14A). Interestingly, we observed most frequently 7-mers containing C and U primarily in the CDS of ER-localized mRNAs. Taking the top 5 identified 7-mers, we derived a HDLBP CU-containing binding motif (Figure 14B).

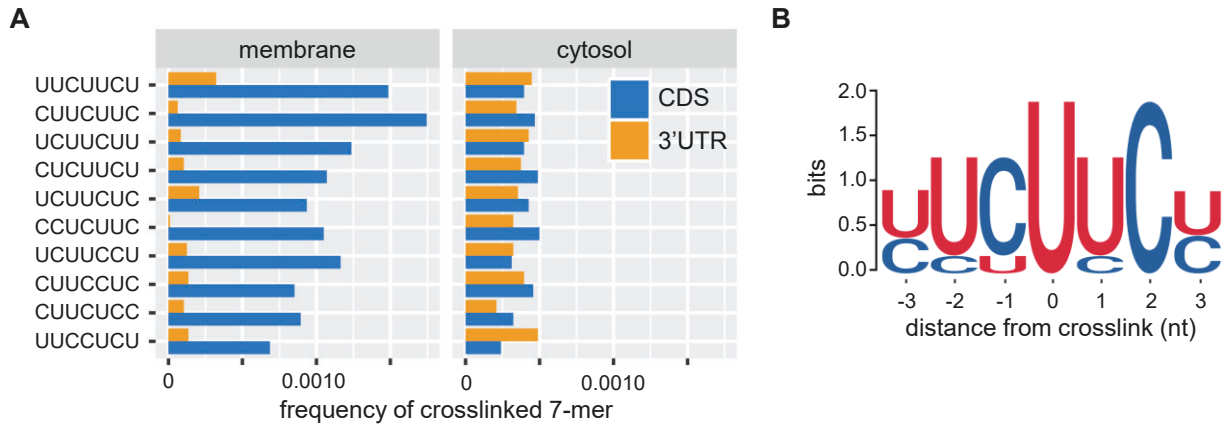

Figure 14 | HDLBP mainly binds to CU-containing sequences.
(A) The frequency of top 10 HDLBP crosslinked 7-mers are shown for membrane- and cytosol-localized mRNAs. Signal detected in the CDS is shown in blue and signal detected in the 3'UTR is shown in orange.
(B) Sequence logo derived from the top 5 HDLBP crosslinked 7-mers.

To understand why HDLBP shows a specificity for CU-containing motifs in the CDS of ER-localized mRNAs, we determined the overall occurrence of these motifs. That clarifies if HDLBP binds the CDS of ER-localized mRNAs because these mRNAs in general contain more frequently CU-motifs. Therefore, we determined all possible 7-mers within the CDS of ER- and cytosol-localized mRNAs. These 7-mers were ranked by their frequency and compared to HDLBP crosslinked 7-mers. Top HDLBP-crosslinked 7-mers were not found among the top 7-mers transcriptome wide, but showed significantly higher occurrence in the CDS of ER-localized mRNAs in comparison to the CDS of cytosol-localized mRNAs (Figure 15). Consequently, the binding preference of HDLBP to CU-containing motifs in the CDS of ER-localized mRNAs can be explained by the sequence composition of ER-localized mRNAs which is different from cytosolic mRNAs.

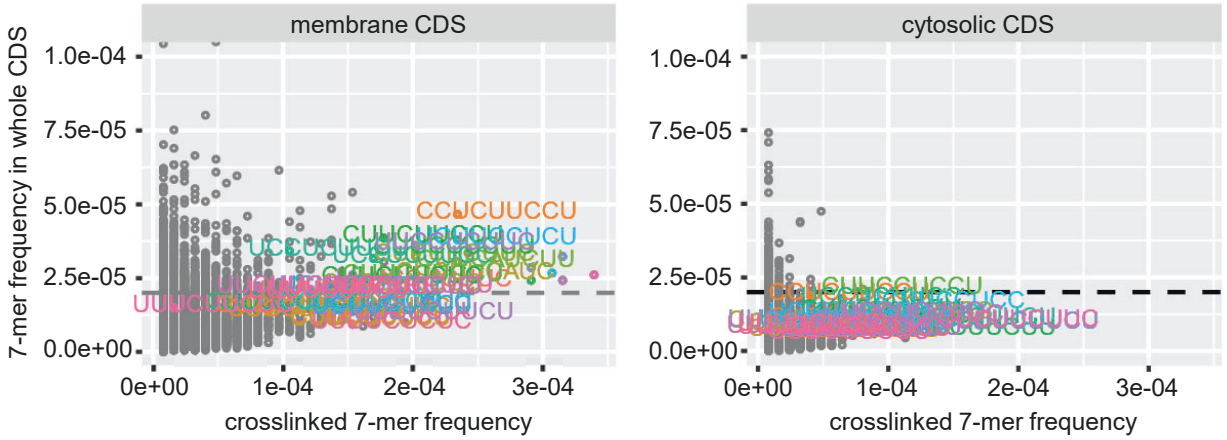

Figure 15 | ER-localized mRNAs contain more frequently HDLBP crosslinked 7-mers. Scatterplots of HDLBP crosslinked 7-mer frequency versus the 7-mer frequency in the whole CDS of membrane-localized mRNAs (left) and cytosol-localized mRNAs (right). The top 20 crosslinked 7-mers are shown in different colors.

Binding of HDLBP to longer RNA recognition elements (RREs)

Furthermore, we investigated if HDLBP could recognize longer RREs. Therefore, we took the top 40 HDLBP k-mers with a length between 4 and 12 nucleotides and compared their frequency in ER-localized versus cytosolic mRNAs (Figure 16A). In general, we identified that crosslinked k-mers were more frequently found in ER-localized mRNAs compared to cytosolic mRNAs. Interestingly, the longer the k-mer the more frequently it was detected in ER-localized versus cytosolic mRNAs. Subsequently, 10-12 nt k-mers showed the greatest difference between ER-localized and cytosolic mRNAs. As a control, we performed the same analysis using all other k-mers and as expected did not observe any difference between ER-localized versus cytosolic mRNAs. Overall, long high affinity RREs for HDLBP binding were significantly more frequently found in ER-localized mRNAs (Figure 16B).

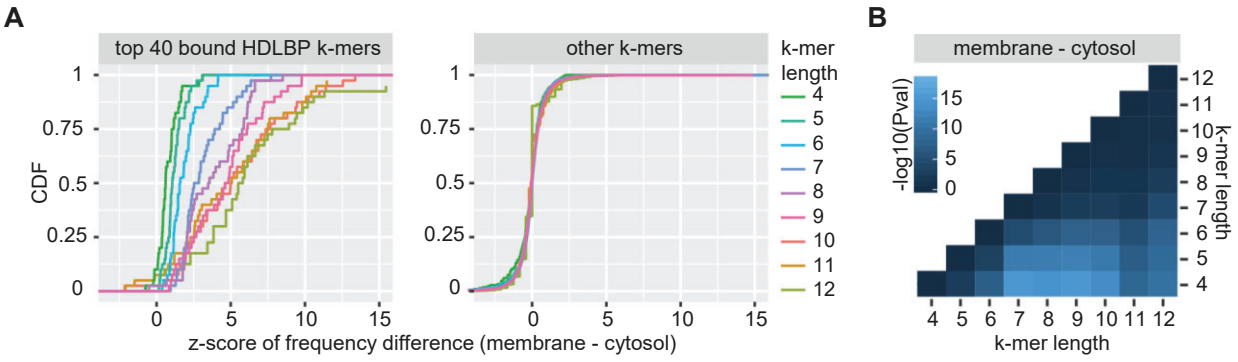

Figure 16 | ER-localized mRNAs contain long high affinity RREs for HDLBP binding. **(A)** Distribution of z-scores derived from differences in the frequency of all possible k-mers within the CDS of membrane-localized mRNAs and the 3'UTR of cytosol-localized mRNAs. This analysis was performed for each k-mer length (4-12) for the group of top 40 bound HDLBP crosslinked k-mers (left) and for all other k-mers (right). **(B)** Log10-transformed p-values of pairwise Wilcoxon rank sum test of z-scores calculated for the top 40 bound HDLBP k-mers as described in (A).

Multivalent interactions of HDLBP

Since HDLBP contains 15 KH domains, we explored the possibility that HDLBP forms multivalent interactions meaning that HDLBP recognizes several RREs within a transcript that are interspaced by unbound nucleotides. Therefore, we investigated a +40 nt / -40 nt region around all detected T-to-C conversions and counted 4-mers. The 4-mers were ranked according to their frequency. We took the top 10 enriched 4-mers and binned the multivalency scores into 5 equally-sized categories (high, mid-high, mid-mid, mid-low or low) and correlated them to the crosslink strength (normalized T-to-C conversions in the +40 nt / -40 nt region) (Figure 17A). Interestingly, we detected that regions with the highest 4-mer frequency showed high crosslinking signal. This suggests that multivalent HDLBP binding was common and gave rise to high affinity interactions.

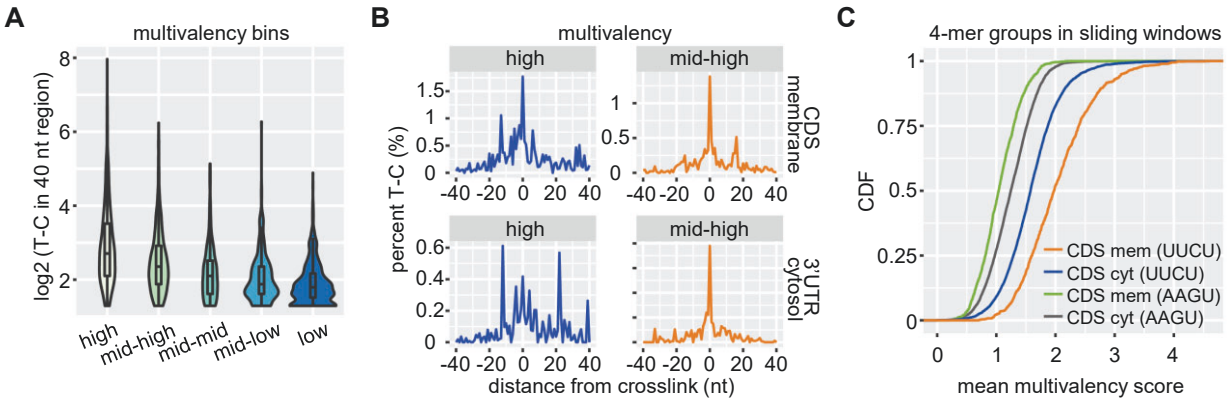

Figure 17 | HDLBP forms multivalent interactions. **(A)** HDLBP multivalency analysis was performed in +40/-40 nt regions around crosslinking sites. For the top 10 enriched 4-mers by frequency the multivalency scores were binned into 5 equally-sized categories and compared to the total normalized T-to-C conversion signal over the +40/-40 nt regions. **(B)** Analysis of the percentage of total T-to-C conversions for every nucleotide position within the +40/-40 nt region for the highest and mid-highest multivalency bins (as defined in (A)). **(C)** Comparison of mean multivalency scores between membrane and cytosol-localized mRNAs in their CDS. A positive set (UUCU, 4-mer group consisting of top10 HDLBP crosslinked 4-mers) and a negative set (AAGU, 4-mer group with no HDLBP enrichment) was used. Occurrence of these 4-mer groups were counted in 30-nt sliding windows. The mean score per transcript was computed and the mean distribution was compared between membrane and cytosol-localized CDS by Wilcoxon rank sum tests.

To further characterize multivalent HDLBP interactions, we investigated in more detail the multivalency bins that we generated. For the high and mid-high multivalency bins, we calculated for each nucleotide within the +40 nt / -40 nt region the percentage of crosslinking signal (total T-to-C conversions) (Figure 17B). Looking at the high multivalency bin, we found a specific crosslink upstream (-13 nt) of the crosslink in the CDS of ER-localized mRNAs and the 3'UTR of cytosolic mRNAs. In addition, we detected a specific crosslink downstream (+16 nt, +20 nt) of the crosslink in the mid-high multivalency bin in the CDS of ER-localized mRNAs and in the high multivalency bin in the 3'UTR of cytosolic mRNAs. Accordingly, we estimated that high affinity HDLBP sites are

approximately 40 nt long and contain 3-4 RREs positioned several nucleotides apart and occur in the CDS of ER-localized mRNAs and the 3'UTR of cytosolic mRNAs.

We further analyzed the frequency of potential HDLBP binding sites in ER-localized and cytosolic mRNAs. Therefore, we calculated within a 30 nt sliding window in the CDS of mRNA sequences the frequencies of specific 4-mers. One group (UUCU) consisted of HDLBP top 10 4-mers and the control group (AAGU) contained 4-mers that showed no HDLBP PAR-CLIP enrichment, but had a similar transcriptome frequency (Figure 17C). We found that the UUCU group was significantly more frequent in the CDS of ER-localized mRNAs compared to cytosolic mRNAs. In contrast, no such significant difference was observed in the AAGU control group. That shows that the CDS of ER-localized mRNAs contains a high density of HDLBP RREs giving rise to multivalent interactions.

Taken together, we identified that HDLBP predominantly binds to a CU-containing motif with a length of up to 12 nt in the CDS of ER-localized mRNAs and forms high affinity multivalent interactions giving rise to an estimated binding site of ~40 nt length containing 3-4 HDLBP RREs. In addition, ER-localized mRNAs and cytosolic mRNAs differ in their sequence composition since we detected in the CDS of ER-localized mRNAs more frequently HDLBP CU-containing binding motifs and overall a high frequency of high density HDLBP RREs.

[6] HDLBP interacts with the translational apparatus

To understand the functional consequences of HDLBP binding to its mRNA targets, we characterized the cellular environment of HDLBP to identify interaction partners. In a previous study, proximity-based proteomics (BioID) of HDLBP and 118 other proteins was performed to uncover RNA regulatory structures (Youn et al., 2018). HDLBP was found to be in close proximity to 849 proteins involved in various cellular processes. In addition, several HDLBP interaction partners have been reported in the past decades of HDLBP research pointing in different directions depending on the proposed function of HDLBP (Cheng and Jansen, 2017). Here, we aim to characterize the cellular environment of HDLBP in HEK293 cells by first analyzing HDLBP RNA interactions and next determining HDLBP protein interactions.

RNA interactions of HDLBP

First, we further explored the PAR-CLIP data for HDLBP RNA interactions. Looking at HDLBP binding to rRNA, we identified 2 major binding sites located in helix 16 and expansion segment 6SB (Figure 18A,B). Interestingly, these sites are in proximity of initiation and elongation factor binding sites which suggests that they have regulatory functions (Anger et al., 2013; Díaz-López et al., 2019).

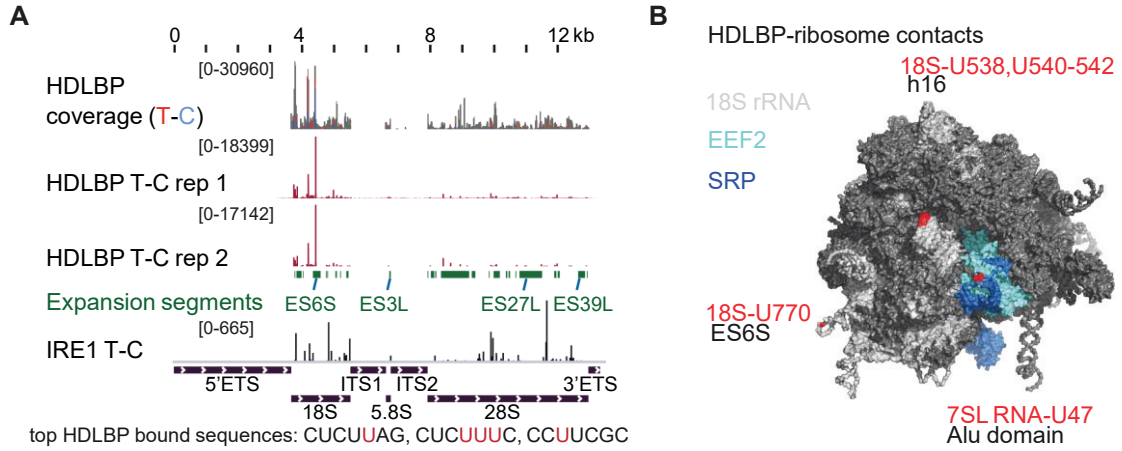

Figure 18 | HDLBP binds to rRNA mainly in helix 16 and expansion segment 6SB. **(A)** Browser representation of HDLBP PAR-CLIP read coverage and T-to-C conversions for pre-ribosomal RNA. PAR-CLIP T-to-C conversions for IRE1 are shown for comparison. **(B)** Structure of the human 80S ribosome (PDB: 4V6X) and the SRP-complex (PDB:3JAJ) were juxtaposed. HDLBP crosslinks to rRNA and 7SL RNA are shown in red.

In addition, HDLBP rRNA binding sites are in close proximity to SRP binding sites. We already showed that PAR-CLIP data suggest that HDLBP is binding to SRP RNA (7SL RNA) and we could confirm this by real-time PCR experiments (see chapter 3). Detailed PAR-CLIP analysis now revealed that HDLBP crosslinks to the small Alu region and to helices of the large (S) domain of 7SL RNA (Figure 19C). To better asses the specificity of HDLBP binding to these sites, we compared it with binding patterns of

other RBPs (IRE1, MOV10) (Figure 19A). We found that HDLBP binding to 7SL RNA is distinct from other RBPs. Furthermore, we compared the percentage of SRP crosslinks over total crosslinks in HDLBP PAR-CLIP data to other RBPs. We found that HDLBP showed similar levels of SRP crosslinks as IRE1 and SSB which are known or expected to interact with the SRP (Figure 19B) (Acosta-Alvear et al., 2018; Garzia et al., 2017; Gogakos et al., 2017). MOV10 is not expected to bind 7SL RNA and therefore showed 10 fold less binding to 7SL RNA compared to HDLBP.

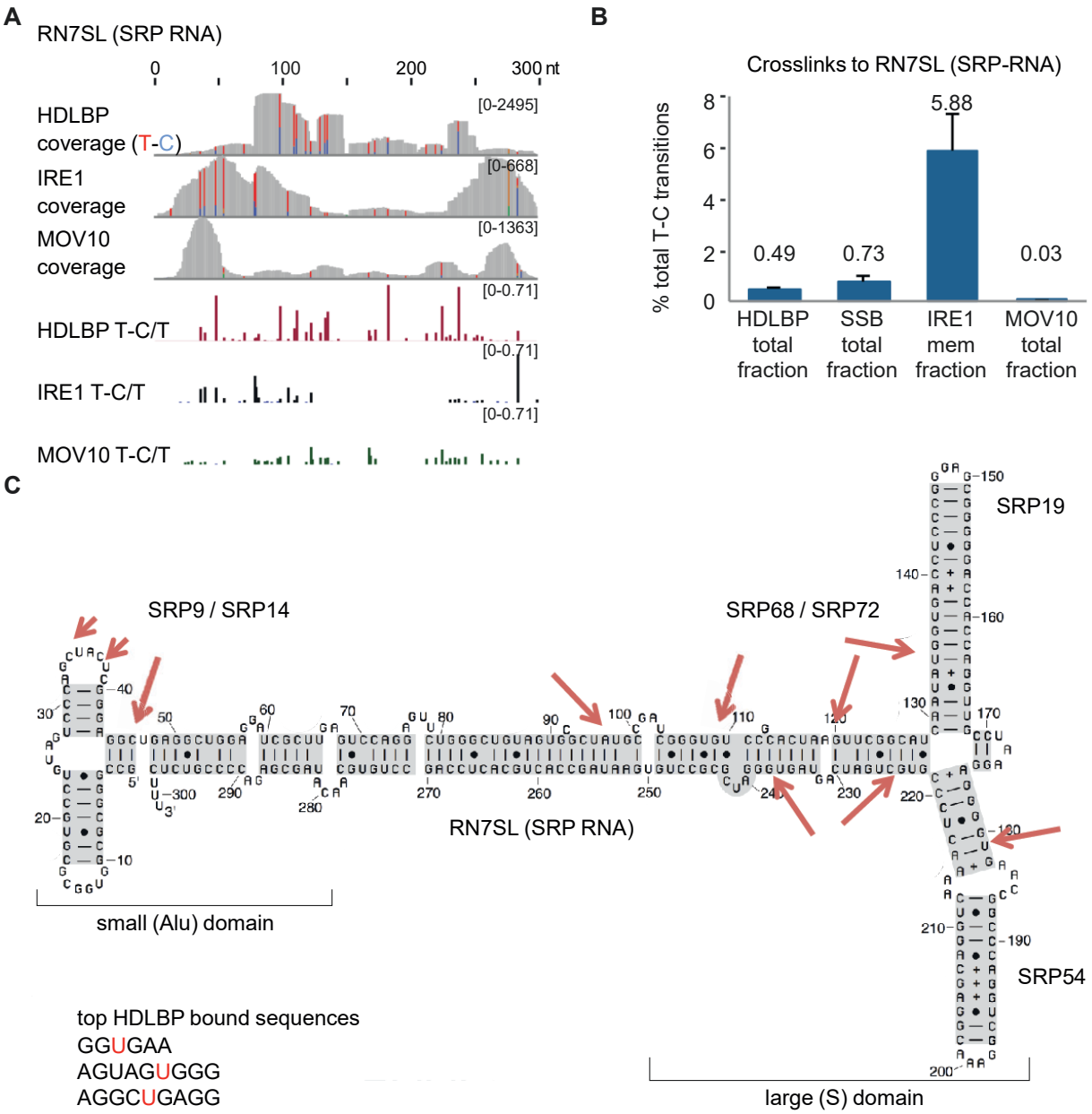

Figure 19 | HDLBP binds to RN7SL (SRP RNA). **(A)** Browser representation of HDLBP PAR-CLIP read coverage and T-to-C conversions for SRP RNA. PAR-CLIP data for IRE1 and MOV10 are shown for comparison. **(B)** Percentage of total T-to-C conversions detected in SRP RNA for HDLBP, SSB, IRE1 and MOV10. **(C)** Secondary structure of SRP RNA. HDLBP crosslinks are indicated with red arrows.

Taken together, PAR-CLIP data suggest that HDLBP interacts with the 40S subunit of the ribosome and the 7SL RNA of the SRP which is known to bind the ribosome in close proximity to the HDLBP ribosome binding sites.

Protein interactions of HDLBP

To further understand the cellular environment of HDLBP, we carried out *in vivo* biotinylation assays (BioID) (Couzens et al., 2013). The experimental outline is shown in Figure 20A. We used a doxycycline inducible cell line that stably expressed HDLBP fused with a biotin ligase at the N-terminus (BirA-FLAG-HDLBP) (see chapter 1, Figure 1). After induction with doxycycline, biotin was added to the cells and all the proteins in close proximity to BirA were biotinylated. Next, those proteins were pulled down and analyzed by mass spectrometry. As a control we used the same protocol but excluded the doxycycline induction step. The experiment was carried out 3 times.

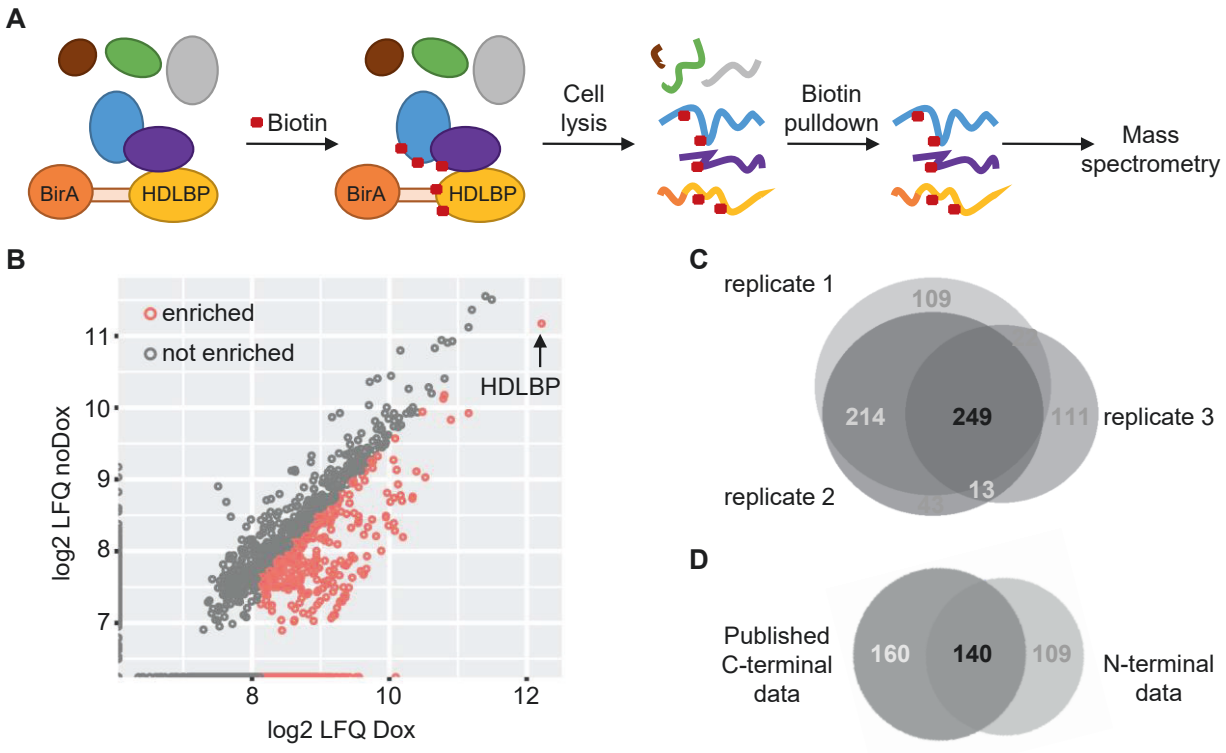

Figure 20 | HDLBP BioID to detect potential protein interactors.
(A) Overview of the BioID experiment. **(B)** Representative scatter plot of log2-transformed LFQ intensity in a doxycycline (Dox)-induced sample versus an un-induced (noDox) control sample. Enriched proteins (LFQ Dox/LFQ noDox>3 and log2(LFQ Dox)>27) are shown in red and the remaining detected proteins are shown in grey. **(C)** Reproducibility of the BioID experiment. The overlap between enriched proteins in 3 replicate experiments is shown in a Venn diagram. **(D)** Venn diagram shows the overlap of proteins in close proximity to HDLBP as detected in this study (N-terminal data) and as detected in the published data set (C-terminal data, (Youn et al., 2018)).

To define proteins in close proximity to HDLBP, we compared the doxycycline induced sample with the control (Figure 20B). Overall, we considered 249 proteins in close proximity to HDLBP since these proteins were present in all 3 replicates (Figure 20C). Comparing our data set to published BioID data obtained from a C-terminal fused biotin ligase HDLBP fusion protein ,we found an overlap of 140 proteins (Figure 20D) (Youn et al., 2018) suggesting that N- and C-terminal data sets correspond well.

Our top 60 N-terminal BioID identified proteins are shown in Figure 21A. They were also identified in the published C-terminal BioID data set (except C1orf198 and VDAC1). Among our top 60 proteins we found a protein of the 40S subunit of the ribosome (RPS3A), a SRP protein (SRP68), proteins involved in translation initiation (EIF4G1, EIF4B, EIF5, EIF4E2), chaperones and chaperonins (HSPA1A, HSPA8, CCT8, HSPA5) and proteins involved in the organization of the cytoskeleton (FLNA, CORO1B, CTTN, FLNB). Top GO terms of the 249 identified proteins included translation and regulation of translation (Figure 21B). In summary, our BioID data suggest that HDLBP is involved in regulation of translation.

To confirm the interaction of HDLBP with the 40S subunit of the ribosome as detected by PAR-CLIP and BioID experiments, we carried out co-immunoprecipitation experiments. Therefore, we used a cell line that stably expressed HDLBP tagged with FLAG/HA (see chapter 1, Figure 1). Anti-Flag along with control anti-IgG immunoprecipitation were carried out and input and eluate samples were analyzed by Western Blot analysis (Figure 21C). We found that HDLBP co-immunoprecipitated with RPS6, a component of the ribosomal 40S subunit of the ribosome. In contrast, only a little amount of a 60S ribosomal subunit component (RPL7) co-immunoprecipitated.

In conclusion, these results suggest that HDLBP is interacting with the 40S ribosomal subunit and the translational apparatus including translation initiation, elongation factors and the SRP.

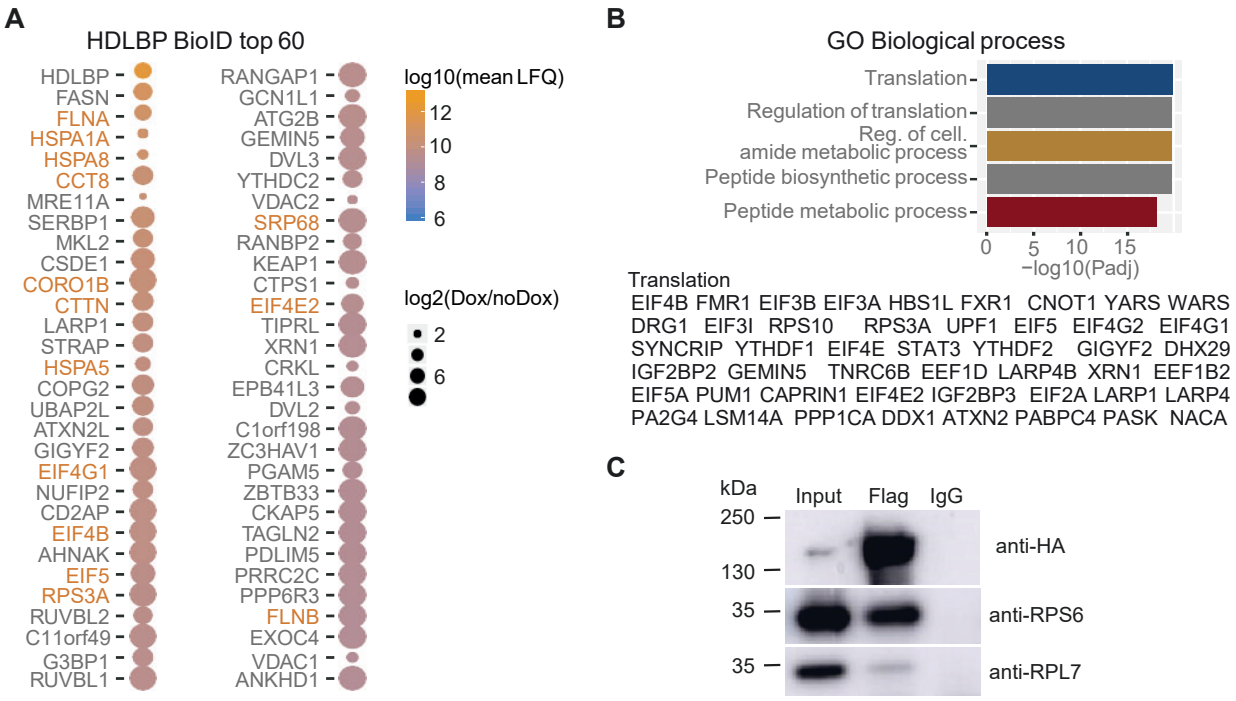

Figure 21 | Proteins in close proximity to HDLBP are involved in translation.
(A) Top 60 enriched proteins of the BioID experiment ranked according to the mean LFQ value from 3 replicates. The color of the dots corresponds to the mean LFQ value and the size of the dots correspond to the enrichment. Protein names labeled in orange are mentioned in the main text.
(B) Gene Ontology enrichment of all enriched BioID proteins. Adjusted P values for the top 5 enriched categories and proteins corresponding to the top category are shown.
(C) Co-immunoprecipitation of FLAG/HA-HDLBP with either anti-FLAG or IgG antibodies. Western Blot analysis of input lysates (0.25%) and eluates (19%) was performed with anti-HA, anti-RPS6 and anti-RPL7 antibodies.

[7] HDLBP does not significantly influence the ER-localization of its mRNA targets

Given the close proximity of HDLBP to proteins that are involved in the organization of the cytoskeleton, we hypothesized that HDLBP might be involved in mRNA transport to the ER. Previously, it was reported that the yeast homolog of HDLBP interacts with polysomes that are transported to the ER (Frey et al., 2001). However, the role of human HDLBP in mRNA transport is unknown.

To uncover the influence of HDLBP on mRNA localization, we carried out cell fractionation experiments in combination with RNA-sequencing. The results of these experiments using HEK293 WT cells are described in detail in Chapter 4. Here, we performed the same experiment using both HDLBP KO cell lines (gRNA1 and gRNA2) and found as expected mRNAs enriched in the cytosol and membrane fractions. Next, we compared the membrane/cytosol ratio of each mRNA between WT and HDLBP KO cells. Assuming that HDLBP transports its targets to the ER, we would expect to encounter less mRNAs in the membrane fraction upon HDLBP KO. Contradictory, we found no significant differences comparing WT to HDLBP KO cells (Figure 22) suggesting that HDLBP is not involved in mRNA transport. Interestingly, one gene, HSPA5, which is also known as BiP, showed less enrichment in the membrane fraction upon HDLBP KO which caught our interest (Figure 22). BiP is an important chaperone and regulator of ER homeostasis. Previously, it was shown that mis-regulation of BiP leads to ER morphology defects (Li et al., 2008). Therefore, we hypothesized that slight mis-regulation of BiP upon HDLBP KO could already cause ER morphology changes and corresponding phenotypes.

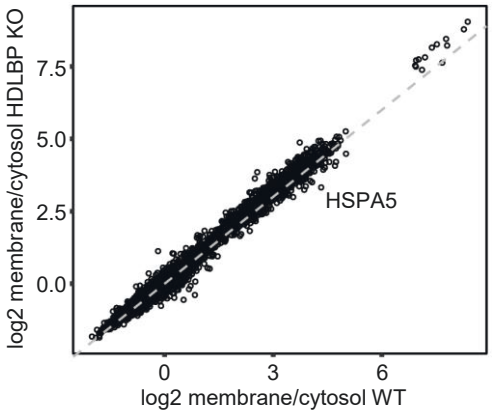

Figure 22 | HDLBP is not influencing the cellular localization of its mRNA targets. Scatterplot of the membrane/cytosol ratio of mRNAs in HEK293 WT vs HDLBP KO cells.

To examine this hypothesis, we carried out electron microscopy and compared the ER morphology of HEK293 WT and HDLBP KO cells. We found that HDLBP KO cells show the same ER morphology as WT cells (Figure 23) suggesting that the slight effect on mis-localization of BiP mRNA upon HDLBP KO does not have consequences on ER morphology.

Overall, our data suggest that HDLBP is not influencing the ER-localization of its mRNA targets.

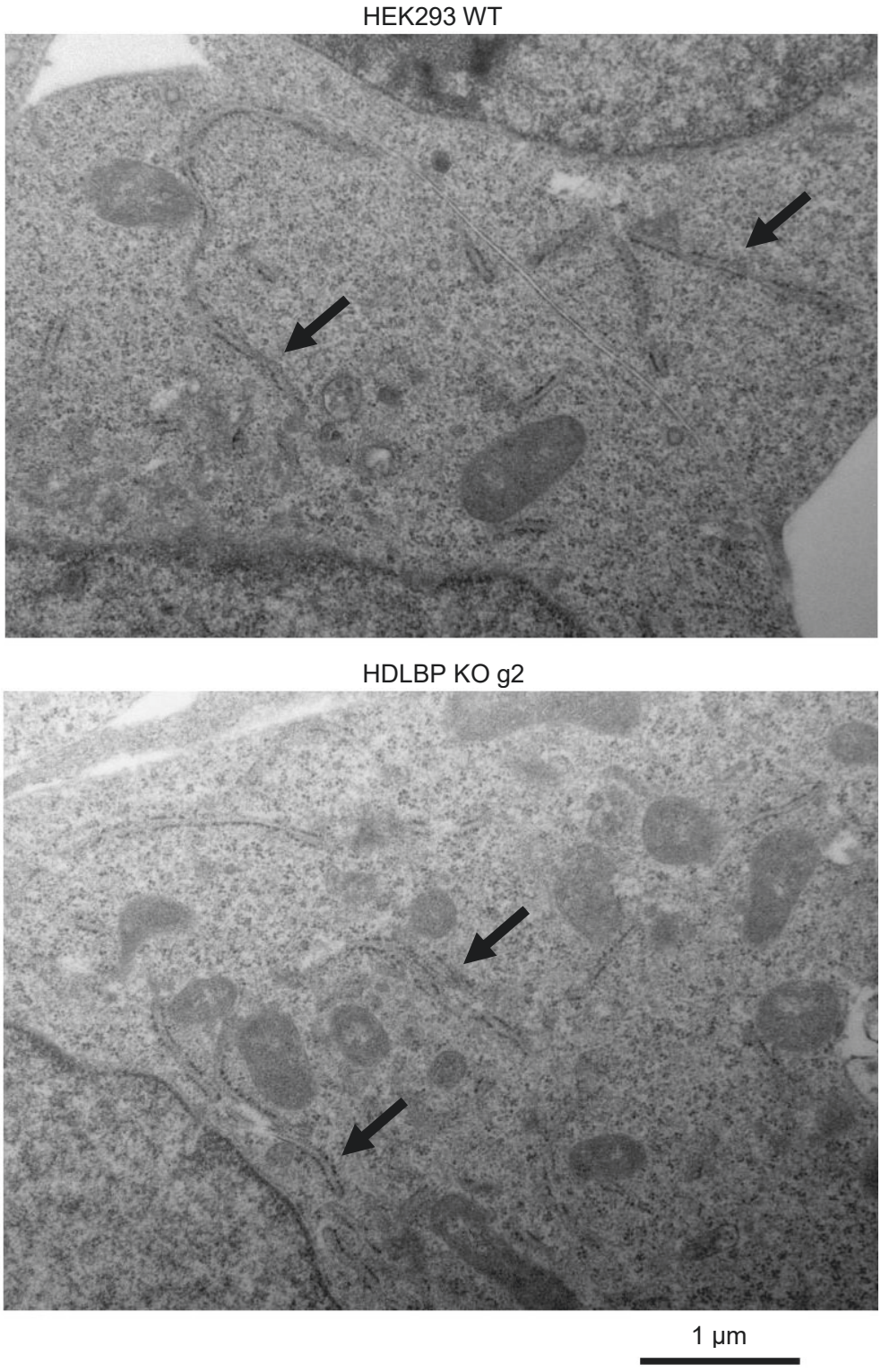

Figure 23 | Representative electron microscopy images of HEK293 WT and HDLBP KO cells showing no ER morphology differences. Arrows indicate the ER.

[8] HDLBP promotes translation of its mRNA targets

Given the interaction of HDLBP with the translational apparatus, we wanted to determine the role of HDLBP in translation. Previously, it was reported that the mouse homolog of HDLBP enhances the translation of a subset of mRNAs encoding secretory proteins (Mobin et al., 2016) and the yeast homolog of HDLBP was also described in several studies to function in translation regulation (Cheng and Jansen, 2017). However, the global effect of human HDLBP on translation remains uncertain. In order to investigate this, we carried out ribosome profiling, pulsed SILAC and polysome profiling.

Ribosome profiling

To determine the role of HDLBP in translation, we performed ribosome profiling (Ingolia et al., 2009) using 2 HEK293 HDLBP KO cell lines (gRNA1 and gRNA2) in comparison to WT HEK293 cells (see chapter 1, Figure 1). The experimental outline is shown in Figure 24A. Cells were lysed with cycloheximide containing lysis buffer and treated with RNase I to enrich ribosome-protected fragments, which were ligated to linkers, converted to cDNA libraries and sequenced. To normalize the data, we also measured the mRNA abundance in the same cell lysates by RNA-sequencing.

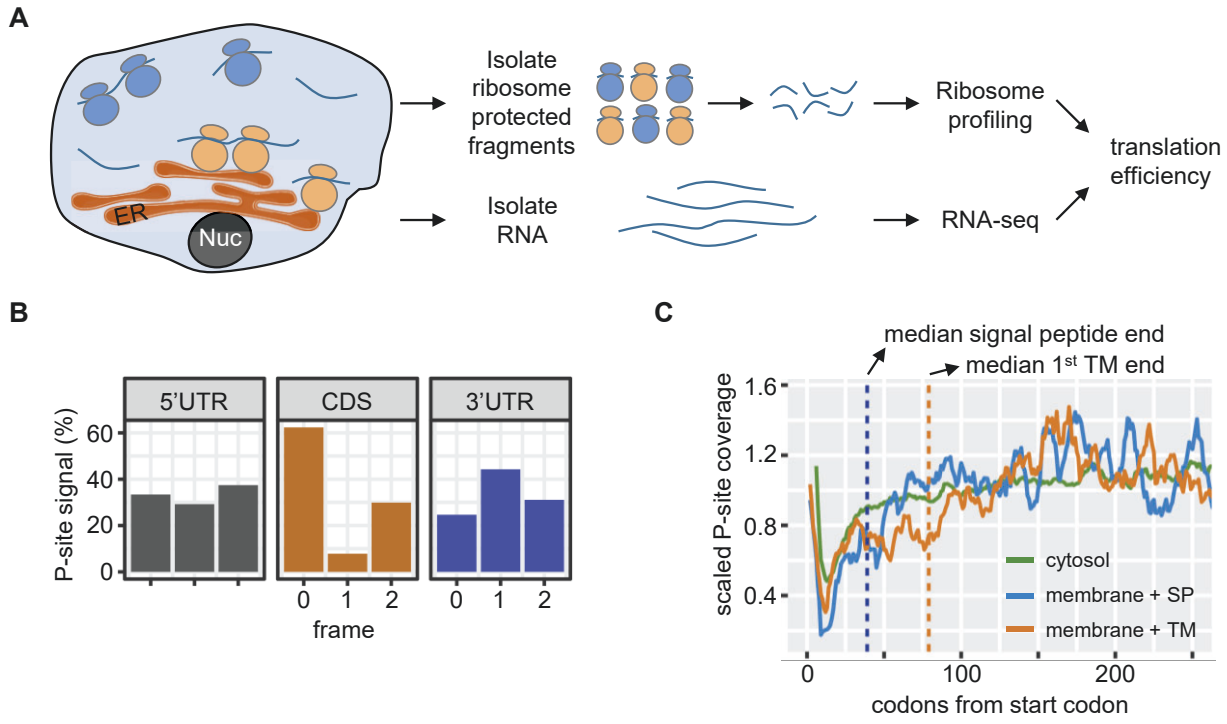

Figure 24 | (A) Overview of the ribosome profiling and RNA-sequencing experiment. **(B)** Representative analysis of in-frame footprints for one ribosome profiling library. The percentage of P sites detected in the 3 possible frames within transcripts is shown in bar plots. **(C)** Scaled P site coverage is shown for 250 codons from the start codon. The scaled P site coverage was averaged over all CDSs and for smoothing a rolling mean over 10 nt was used. Signal from cytosolic mRNAs is shown in green, membrane-localized mRNAs containing a signal peptide (SP) or a transmembrane domain (TM) are shown in blue and orange respectively.

Ribosome profiling was carried out in replicates and after processing the reads, we identified as expected ribosome footprints enriched in the CDS showing a typical codon periodicity pattern (Figure 24B). Considering that we are especially interested to uncover the effect of HDLBP on its mRNA targets, which are mainly ER-localized, we asked if our ribosome profiling data set from non-fractionated cells sufficiently captured ER-localized mRNAs. Therefore, we checked the footprint density of SP and TM containing mRNAs from the start codon to the region where the targeting signal (SP or 1st TM) emerged from the ribosome. We observed as expected low footprint density in that region confirming that we are able to capture the translation of ER-localized mRNAs with near-nucleotide resolution (Figure 24C).

To assess the influence of HDLBP on its ER-localized mRNA targets, we calculated the translation efficiency (TE) changes comparing HDLBP KO to WT (Figure 25A). Evaluating the TE changes has the advantage that both mRNA abundance and ribosome occupancy are taken into account. Furthermore, we defined for this analysis different similar sized groups of ER-localized mRNAs. Groups were composed of mRNAs with high, mid, low or no HDLBP crosslinks in the CDS as determined by PAR-CLIP. Strikingly, we found that strong HDLBP mRNA targets (having high numbers of T-to-C conversions in the CDS) showed the highest decrease in TE upon HDLBP KO. We observed for mRNAs having mid numbers of HDLBP PAR-CLIP T-to-C conversions a modest decrease in TE upon HDLBP KO and we found no TE differences for low HDLBP mRNA targets and HDLBP non-targets. These results demonstrate that HDLBP promotes the translation of its target mRNAs.

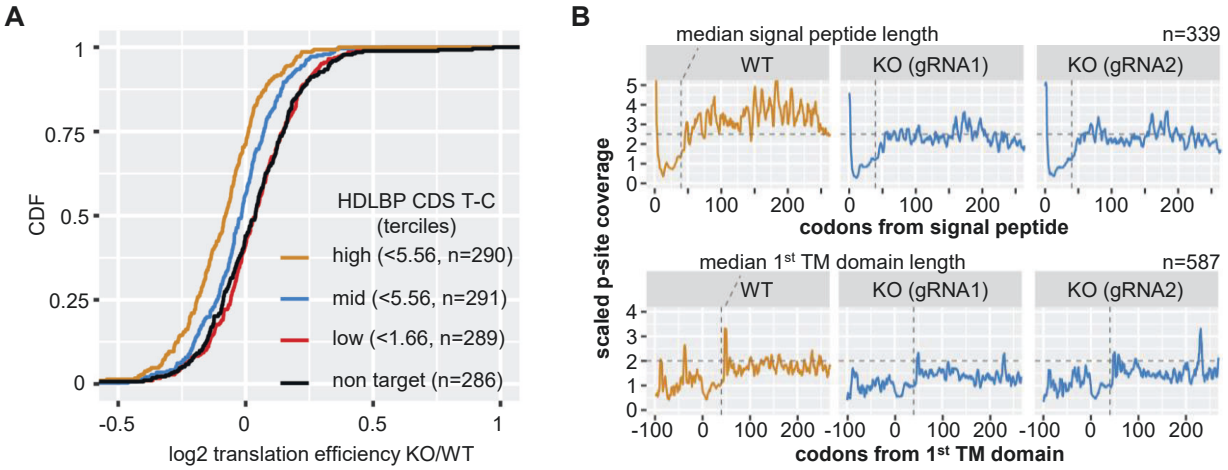

Figure 25 | HDLBP promotes the translation of its mRNA targets.
(A) Differences in translation efficiency between HDLBP KO to WT cells are shown for different groups of ER-localized HDLBP targets. To test for significance, the Wilcoxon rank sum test was used.
(B) Ribosome P site coverage around signal peptides (upper panel) and the 1st transmembrane domain (lower panel) was compared between HDLBP KO and WT cells. P site coverage was scaled to the coverage in codons 20-40 of each mRNA and a rolling mean of 5 nt was used to smooth the profiles. Median signal peptide and 1st transmembrane domain length are indicated with a vertical dotted line. Absolute numbers of analyzed mRNAs are shown.

To further understand the interaction of HDLBP with its mRNA targets, we analyzed the footprint coverage around known targeting signals (Figure 25B). Therefore, we aligned mRNAs according to their targeting signal. The upper panel shows the average P site coverage 250 codons after the start of the signal peptide of 339 signal peptide encoding mRNAs. The lower panel shows the average P site coverage 100 codons before and 250 codons after the start of the first transmembrane domain of 587 mRNAs that encode at least one transmembrane domain. Interestingly, we found that the ribosome density downstream of the targeting signal region (SP or the 1st TM) is decreased upon HDLBP KO suggesting that HDLBP contributes to the translation arrest.

pulsed SILAC

The ribosome profiling experiments showed that the TE of HDLBP targets is decreased upon HDLBP KO suggesting that less proteins are synthesized. To confirm this, we carried out pulsed stable isotype labelling with amino acids in cell culture (pSILAC) (Schwanhäusser et al., 2009) in combination with cell fractionation to obtain the cytosol and membrane fraction. The experimental outline is shown in Figure 26. HDLBP KO and WT cells were grown in medium containing the "light" form of amino acids. Next, this medium was replaced by either medium containing the "heavy" form or the "medium" form of amino acids for 2 or 4 hours. Cells were fractionated as described in Chapter 4 and thereby WT and HDLBP KO cells (labeled with different forms of amino acids) were combined.

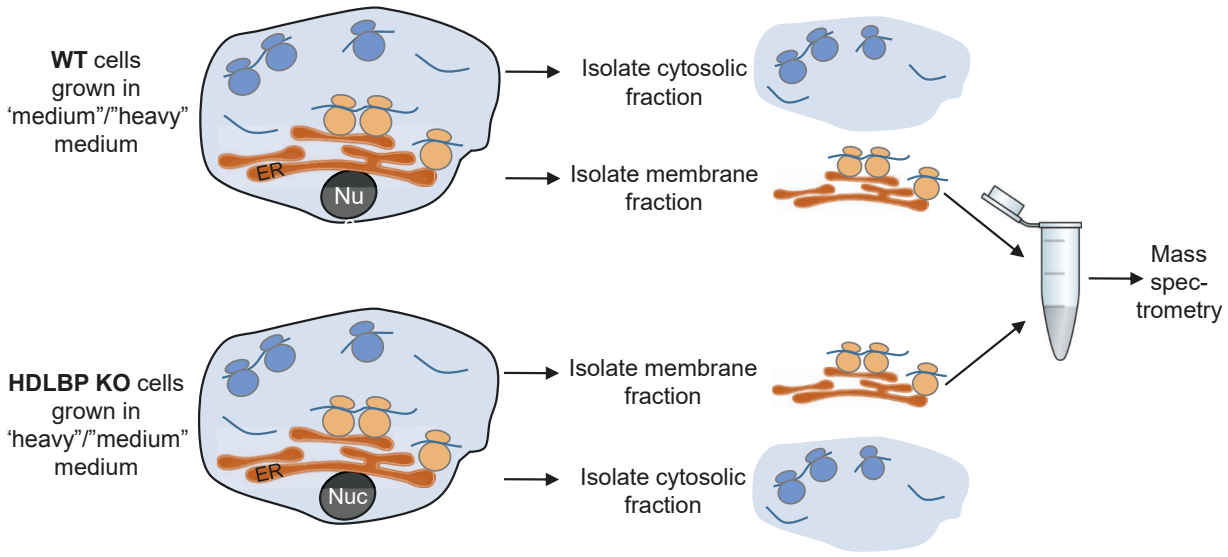

Figure 26 | Overview of the pSILAC experiment.

The heavy/medium ratios which recapitulate the newly synthesized proteins in HDLBP KO vs WT were calculated and proteins were grouped according to their mRNA localization and HDLBP binding (Figure 27A). We found a decrease in protein synthesis in the membrane fraction for proteins encoded by ER-localized mRNAs. Looking at proteins encoded by mRNAs that are strong HDLBP targets (based on the amount of T-to-C conversions in the CDS), we found as expected a reduction of protein synthesis (Figure 27B). The top 30 proteins with most decreased synthesis are shown in Figure 27C.

Taken together, HDLBP's interaction with its target mRNAs influences their efficient protein synthesis.

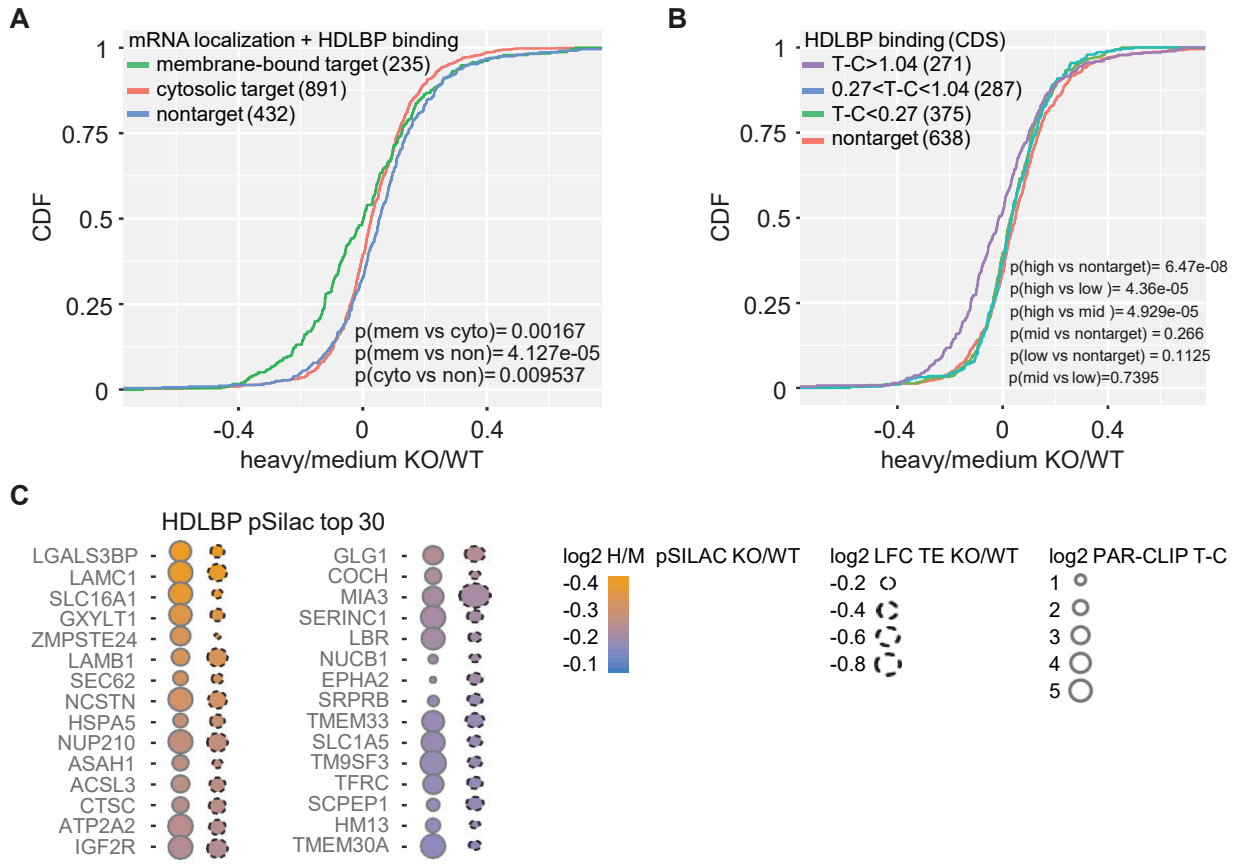

Figure 27 | Proteins encoded by ER-localized mRNAs show decreased protein synthesis upon HDLBP KO.
(A+B) pSilac analysis of newly synthesized proteins in the membrane fraction of WT and HDLBP KO cells. SILAC heavy vs. medium ratios (H/M) were quantified and reflect changes in protein synthesis upon HDLBP KO. Wilcoxon rank sum test was used to compare SILAC ratios between different groups of HDLBP targets.
(C) The top 30 proteins with most decreases in protein synthesis upon HDLBP KO as detected by pSILAC are shown. The sizes of the dots correspond to PAR-CLIP T-to-C conversion signal (left dot with solid outline) or to decrease in translation efficiency upon KO as detected by ribosome profiling (right dot dashed outline).

Polysome profiling

To further define the role of HDLBP in translation, we carried out polysome profiling (Jagannathan et al., 2011). This technique is based on sucrose density centrifugation of the cell lysate which allows to separate free mRNAs, the 40S and 60S subunits of the ribosomes, the monosomes and the polysomes. Subsequent Western blotting of the separated fractions allows to visualize the distribution of the protein of interest. Previously, the yeast homolog of HDLBP was found to be enriched in the polysome fraction suggesting that its involved in translation regulation (Baum et al., 2004; Brykailo et al., 2007; Frey et al., 2001; Hirschmann et al., 2014; Lang and Fridovich-Keil, 2000; Lang et al., 2001; Li et al., 2004).

Here, we performed polysome profiling using a cell line that stably expressed HDLBP tagged with FLAG/HA (see chapter 1, Figure 1). After recording of the profile and collection of the fractions, we carried out TCA precipitation and Western Blot analysis (Figure 28). We found that HDLBP is enriched in fraction 1-6. In addition, we also detected HDLBP in the monosome fraction and in the polysome fractions 8-12. As positive controls, we used the ribosomal proteins RPS6 and RPL7 and as a negative control we used tubulin. All controls were distributed as expected.

We conclude that human HDLBP mainly interacts with the ribosomal subunits, whereas the yeast homolog of HDLBP was mainly detected in the polysome fractions.

Taken together, we show that HDLBP promotes the translation of its mRNA targets resulting in their efficient protein synthesis. Furthermore, we find evidence that HDLBP contributes to translation arrest. Polysome profiling experiments support the role of HDLBP in translation.

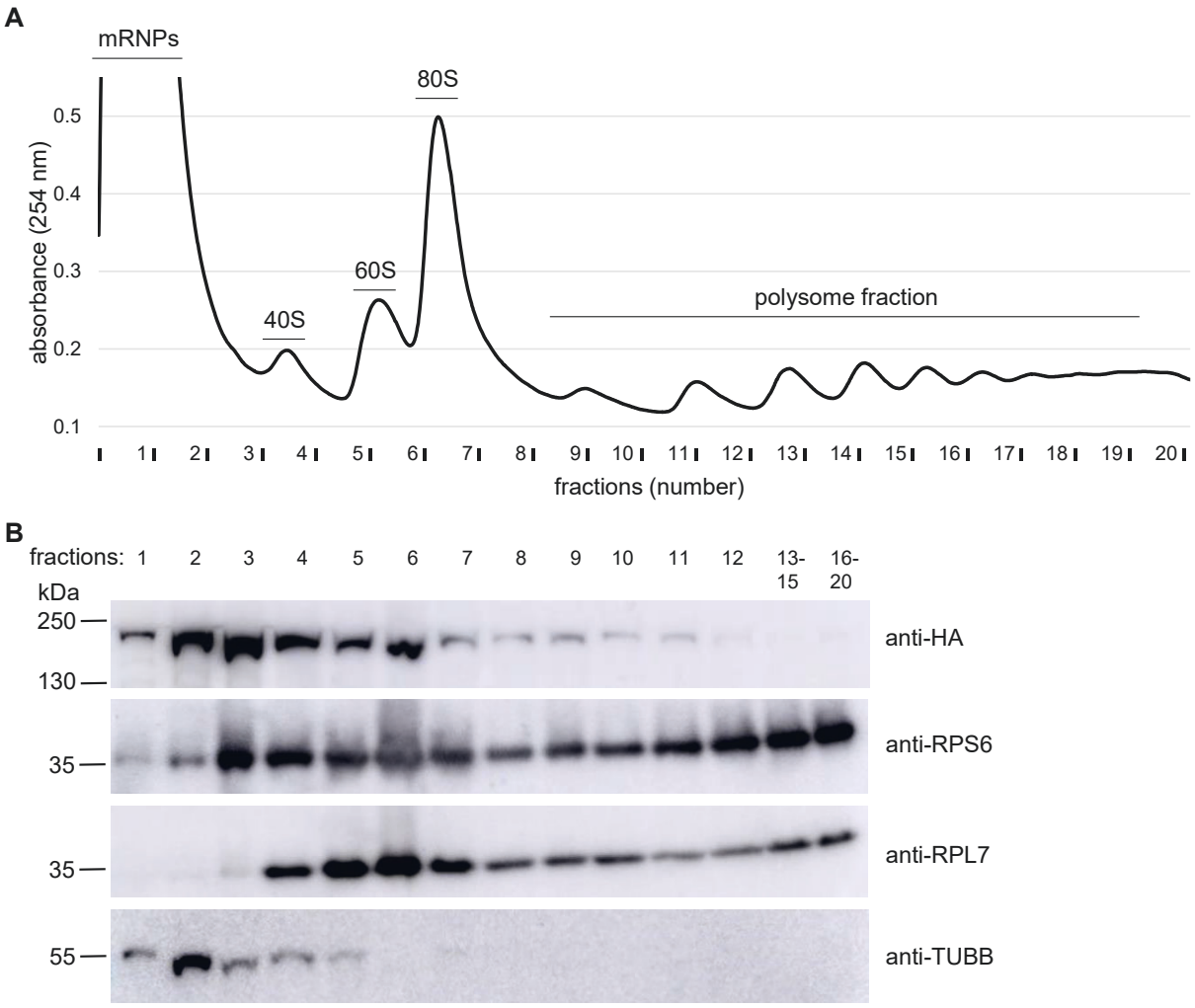

Figure 28 | Polysome profiling shows that HDLBP is enriched in mRNPs and ribosomal subunits, but also associated with monosomes and polysomes.
(A) Representative polysome profiling of FLAG/HA-HDLBP HEK293 cells.
(B) Representative Western Blot analysis with TCA precipitated polysome profiling fractions. For fraction 3-12 all the material was used, whereas only 20% of fraction 1 and 2 were loaded. Fractions 13-15 and 16-20 were combined and an equal volume was loaded. A primary anti-HA antibody was used to detect FLAG/HA-HDLBP. Antibodies against RPS6, RPL7 were used as positive controls and an anti-tubulin (TUBB) antibody was used as negative control.

[9] HDLBP prevents global stalling

To further characterize the impact of HDLBP on translation, we analyzed the ribosome occupancy per codon. During translation each nucleotide of a codon is bound by a ribosome at a designated site (A, P and E sites). Our ribosome profiling data allowed us to extract P site and E site coverage for each codon. Next, we compared the P site and E site occupancy between HEK293 WT and HDLBP KO cells (Figure 29). We found that various codons showed higher ribosomal occupancy in both P site and E site upon HDLBP KO suggesting that HDLBP absence leads to global stalling. Interestingly, highest ribosome occupied codons encoded for Phenylalanine (UUC, UUU) and Leucine (CUU). Those codons correspond to the identified HDLBP CU-containing binding motif (see chapter 5).

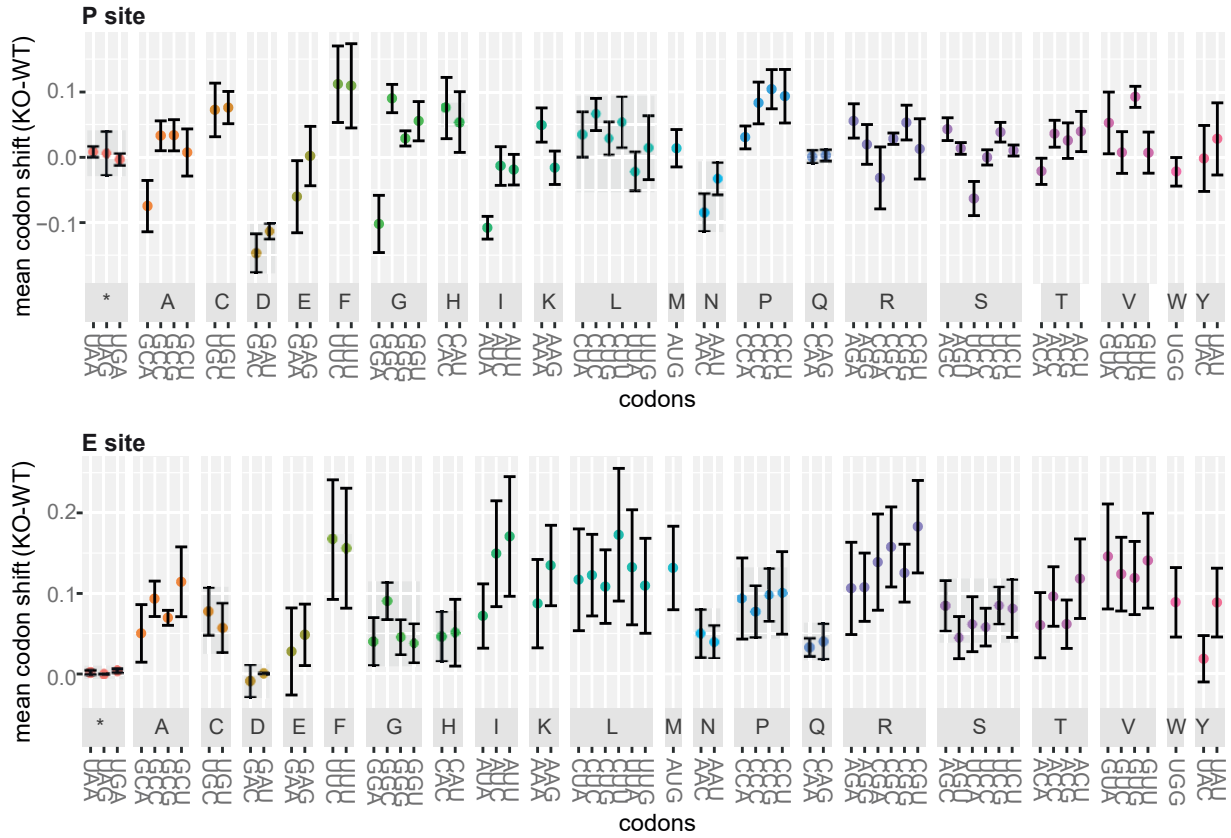

Figure 29 | HDLBP KO leads to higher ribosome occupancy of various codons in both P Site and E site. The difference in codon frequencies in the P site (top) and E site codons (bottom) in HDLBP KO vs. WT are shown for each codon. Mean codon shift was calculated for four replicates and the mean standard deviation is shown by error bars.

To understand if HDLBP is indeed binding to these codons (UUC, UUU, CUU), we investigated the PAR-CLIP signal on codon level. This will clarify which codons are bound by HDLBP because the CU-containing motif alone gives rise to a number of possible codons that could be bound by HDLBP. We counted T-to-C conversions for each codon per transcript, normalized to the transcript expression level and sorted the transcripts according to their membrane enrichment as determined by cell fractionation and RNA-sequencing (see chapter 4). We found that HDLBP most bound codons are UUC encoding for phenylalanine and CUC, CUU encoding for leucine for which we observed high ribosome occupancy upon HDLBP absence (Figure 30). Therefore, HDLBP binding on these codons might prevent stalling.

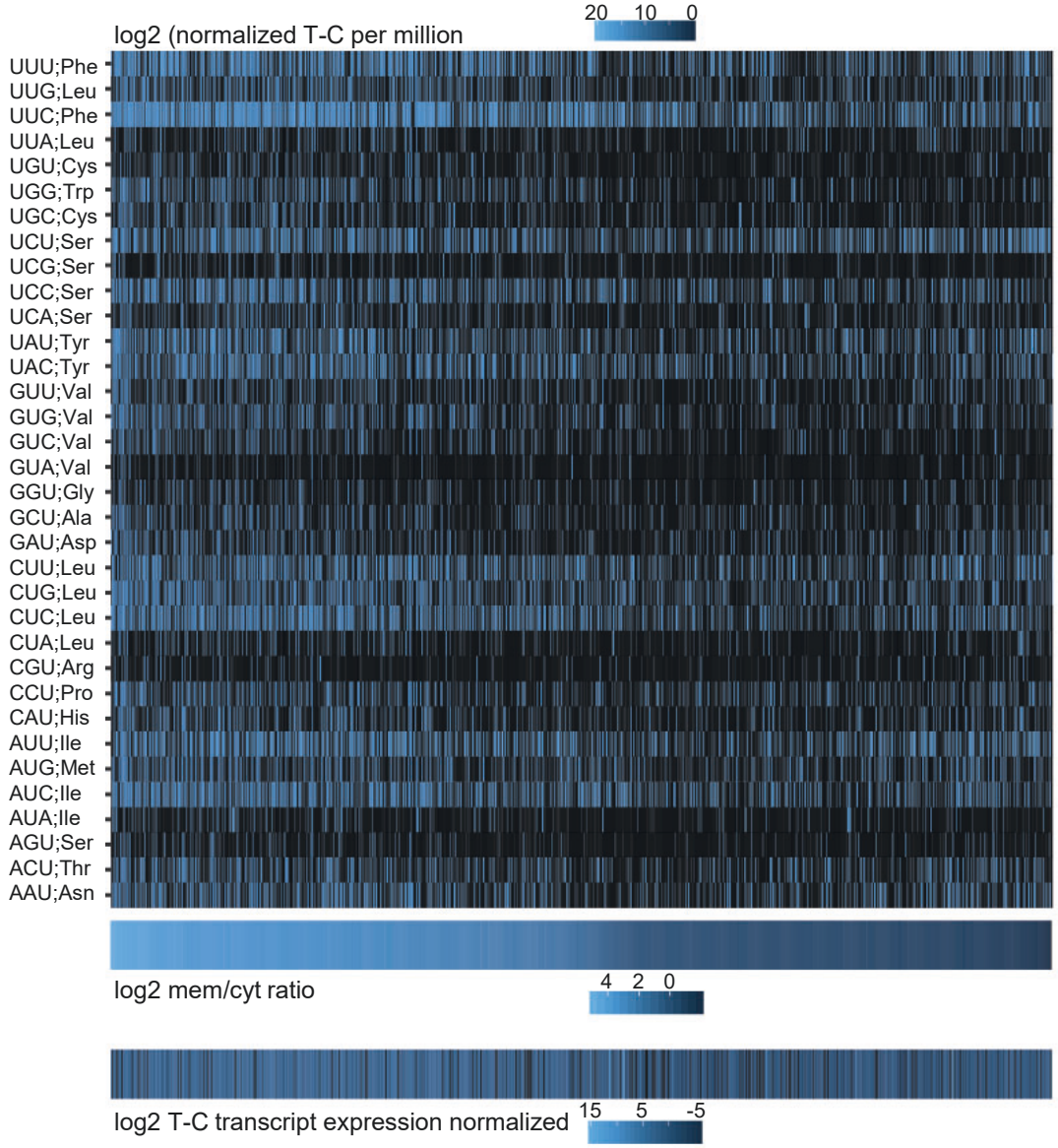

Figure 30 | HDLBP mostly binds CU-containing codons. The normalized T-to-C conversions were counted for each codon per transcript and normalized to the transcript expression level. The heatmap shows transcripts (columns) ranked according to their membrane enrichment as determined by cell fractionation and RNA-sequencing.

Ribosome stalling can be caused by mis-regulation of tRNA availability. We already identified codons that show stalling and therefore we can estimate which tRNAs could be less available. To understand the impact of HDLBP on those tRNAs, we further explored the PAR-CLIP data and characterized HDLBP binding to tRNAs. We counted the PAR-CLIP coverage for each tRNA and also investigated the position of T-to-C conversions within tRNAs (Figure 31). Interestingly, 4 Leucine isotype tRNAs were among the top 15 enriched tRNAs and we detected the crosslinking sites in the variable and D loops. The crosslinking sites corresponded to the identified HDLBP CU-containing motif (UCUUC). Codons decoded by these tRNAs showed high ribosomal occupancy upon HDLBP KO suggesting that HDLBP tRNA binding contributes to efficient tRNA decoding.

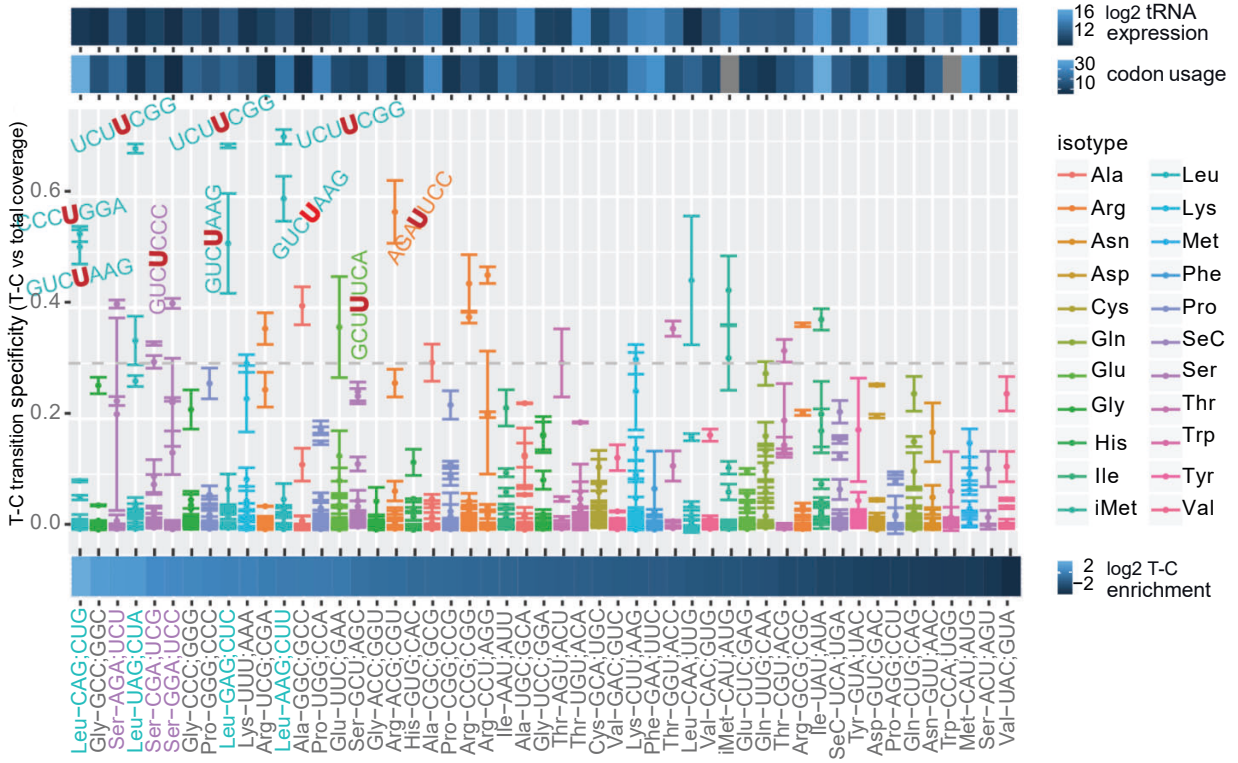

Figure 31 | HDLBP binding to tRNAs. HDLBP PAR-CLIP detected T-to-C conversions in tRNAs were normalized to total tRNA abundance and ranked from the highest to the lowest value (left to right). For each T-to-C conversion the conversion specificity (T-C transition vs. total read coverage) was displayed. On top the total log2-transformed tRNA abundance and codon usage are shown.

In summary, codon level analysis of ribosome profiling and PAR-CLIP data revealed that HDLBP KO leads to higher ribosomal occupancy for various HDLBP bound codons suggesting that HDLBP prevents global stalling. Moreover, we identified that HDLBP binds tRNAs that decode these codons suggesting that HDLBP contributes to efficient tRNA decoding.

[10] HDLBP promotes secretion and transmembrane protein abundance

Since HDLBP promotes the translation and synthesis of several secretory and membrane proteins, we wanted to determine the effect of HDLBP on secretion and membrane insertion per se. Previously, it was shown that HDLBP influences the secretion of certain proteins from primary mouse hepatocytes (Mobin et al., 2016). However, the effect of human HDLBP on secretion and in non-specialized secretory cells is unknown. Moreover, the impact of HDLBP on membrane insertion is unexplored.

The influence of HDLBP on secretion

To study secretion in HEK293, we used reporter proteins because a global mass spectrometry based assessment of all proteins secreted from HEK293 cells resulted in non-quantifiable results (data not shown). As reporter proteins, we used secreted *Gaussia* luciferase (Gluc) and secreted alkaline phosphatase (SEAP). The experimental outline is shown in Figure 32. Gluc and SEAP were transiently transfected into HEK293 WT and HDLBP KO cells. The enzyme activity was quantified in culture medium. In order to normalize for transfection efficiency and cell number differences, we co-transfected non-secreted Firefly luciferase (Fluc) and quantified intracellular Fluc activity.

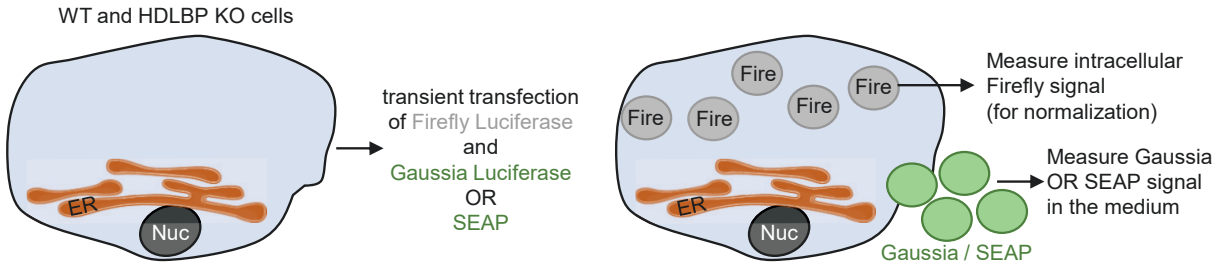

Figure 32 | Overview of the Gluc/SEAP luciferase secretion assay.

Next, we calculated the Gluc/Fluc and SEAP/Fluc ratio for WT and HDLBP KO cells. Gluc and SEAP activity were significantly decreased around 30% upon HDLBP KO (Figure 33A,B) showing that HDLBP promotes secretion.

To prove that the reduced secretion effect in HDLBP KO cells was due to HDLBP absence, we performed rescue experiments. Therefore, we used HDLBP KO gRNA2 cells because here, we observed a slightly stronger secretion reduction in comparison to HDLBP KO gRNA1 cells. Moreover, SEAP was secreted less compared to Gluc in HDLBP KO gRNA2 and consequently we performed the SEAP assay using HDLBP KO gRNA2 cells in comparison to WT cells. The experiment was carried out as described above but in addition, we co-transfected HDLBP. We found that co-transfection of

HDLBP resulted in more SEAP secretion in both WT and HDLBP KO cells (Figure 33C). Thus, we successfully rescued the phenotype and further demonstrate that HDLBP promotes secretion.

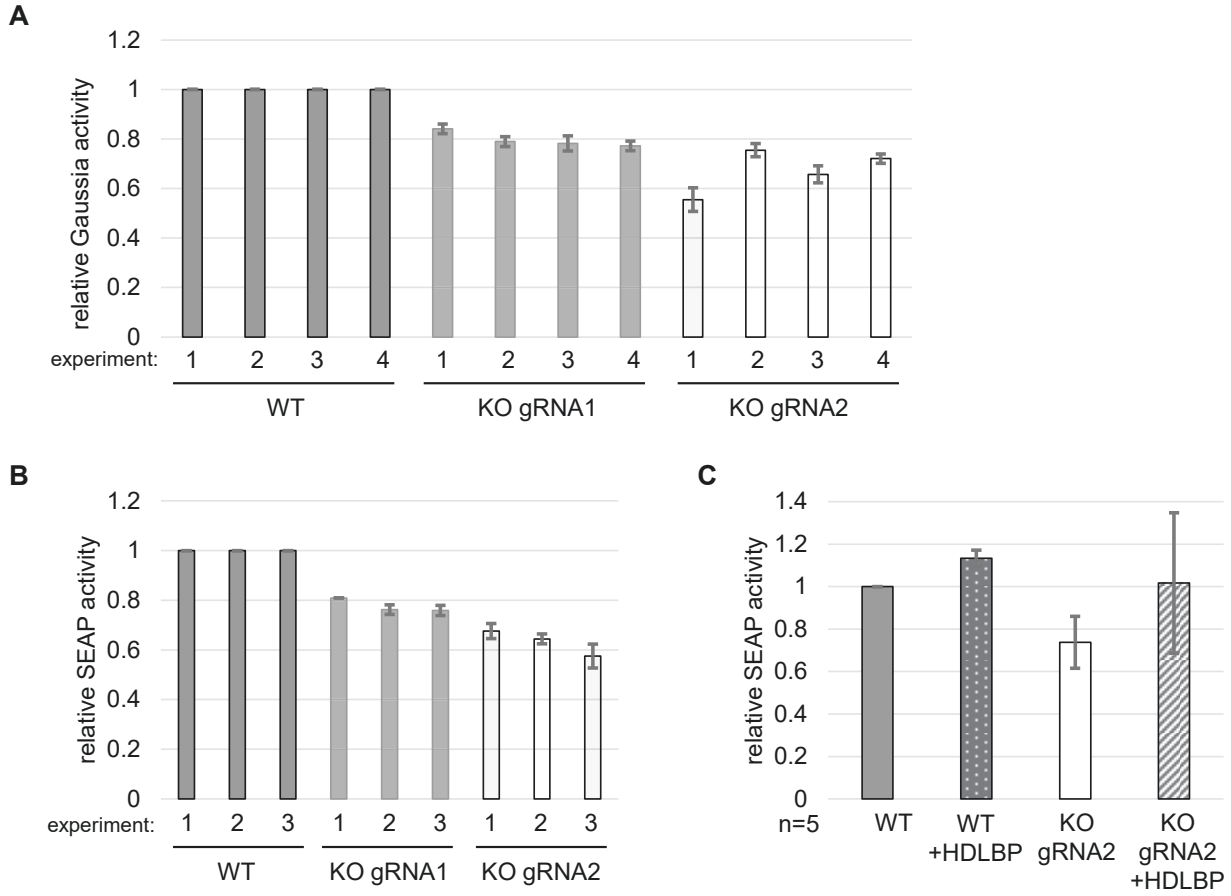

Figure 33 | HDLBP promotes secretion of SEAP and Gluc.
(A+B) Gluc activity (A) or SEAP activity (B) was quantified in cell culture medium and normalized to the intracellular Fluc activity. The average of 5 technical replicates per experiment is shown and the corresponding standard deviation as indicated by the error bar.
(C) SEAP activity was quantified in cell culture medium and normalized to the intracellular Firefly luciferase activity. The average value of 5 SEAP rescue experiments with co-transfection of HDLBP is shown and the corresponding standard deviation as indicated by the error bar. For each of the 5 experiments at least 5 technical replicates were used.

The influence of HDLBP on membrane-protein surface expression

To study the influence of HDLBP on membrane insertion, we chose to determine the surface expression of one HDLBP target TFRC also known as CD71. We already showed by PAR-CLIP and real-time PCR that TFRC mRNA is bound by HDLBP (see chapter 3). In addition, ribosome profiling showed that TFRC has a decreased translation efficiency upon HDLBP KO and pSILAC confirmed that less TFRC protein is synthesized in absence of HDLBP (see chapter 8). To quantify surface expression of TFRC, we stained WT and HDLBP KO cells with a TFRC flow cytometry suitable antibody and measured the mean fluorescent intensity (MFI) of the antibody signal by flow cytometry (Figure 34A,B). We found that the relative MFI was significantly decreased around 25% upon HDLBP KO, whereas our control staining showed as expected no difference between HDLBP KO and WT.

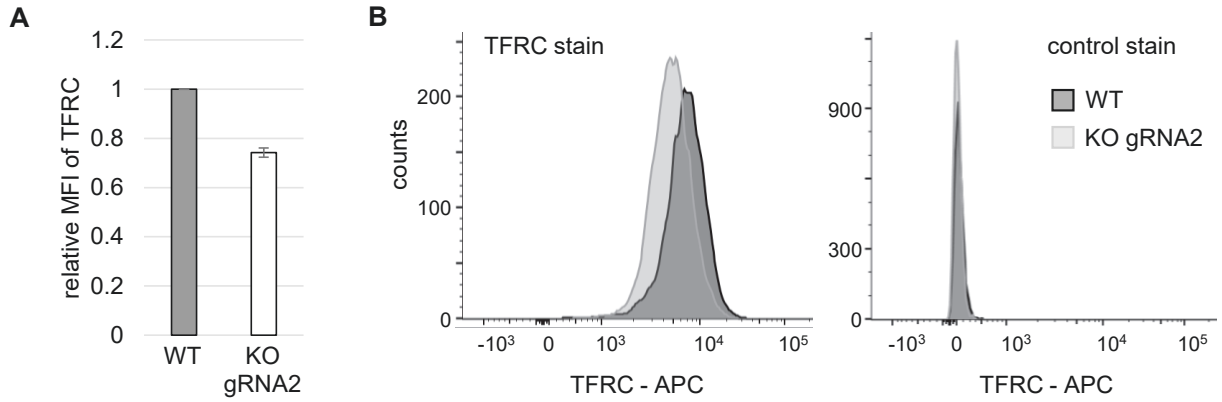

Figure 34 | TFRC shows less surface expression upon HDLBP KO.
(A) Relative mean fluorescent intensity (MFI) of TFRC in WT and HDLBP KO cells as determined by flow cytometry experiments. The average of 4 replicates is shown and the corresponding standard deviation as indicated by the error bar.
(B) Representative histograms of TFRC-APC (left) and an isotype-control (right) staining of WT and HDLBP KO cells.

Overall, HDLBP influences the surface expression of membrane proteins whose mRNA it binds and promotes the secretion of reporter proteins.

[11] HDLBP decreases tumor formation capacity *in vivo*

To understand the impact of HDLBP on cancer biology, we established a HDLBP KO cell line in lung carcinoma cells A549 (see chapter 1, Figure 1). First, we analyzed the 2D growth characteristics of these cells in comparison to WT A549 cells. We found that HDLBP KO cells grew slower and were less viable (Figure 35A). Scratch migration assays showed that HDLBP KO cells had a reduced wound closure capacity (Figure 35B).

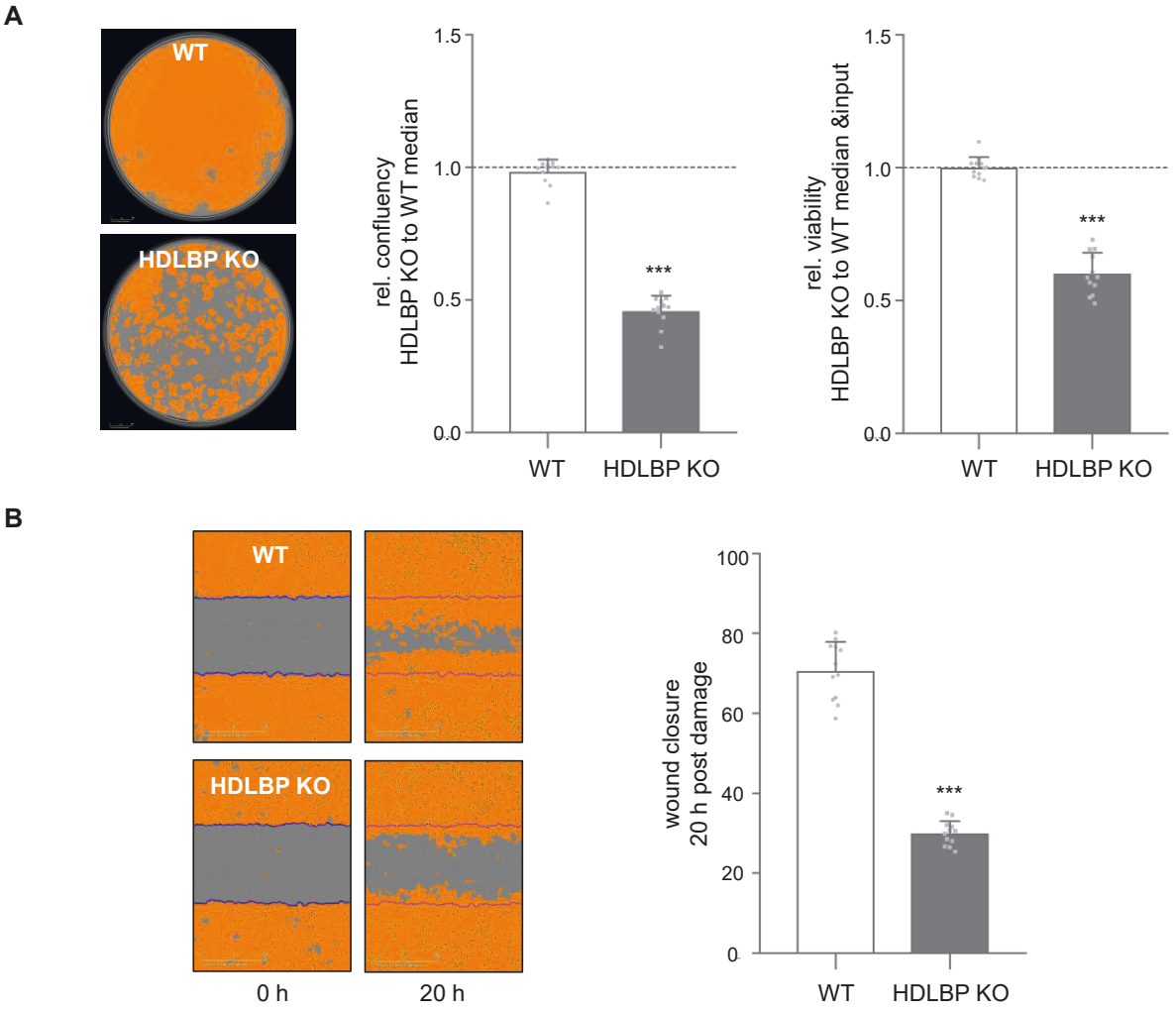

Figure 35 | A549 HDLBP KO cells grow slower and are less viable compared to WT cells.
(A) Relative confluence and viability of A549 WT or HDLBP KO cells was quantified and the left panel shows representative confluence masks.
(B) Wound healing of A549 WT or HDLBP KO cells was quantified after 20 h post damage and the left panel shows representative confluence masks.

Next, we tested the tumor formation capacity of A549 HDLBP KO and WT cells *in vivo*. To this end, cells were stably transduced with infrared fluorescent protein (iRFP) and injected into 8 athymic nude mice. Tumor initiation and growth was measured 0, 7, 14 and 21 days post-injection (Figure 36A,B). WT cells formed as expected tumors in all 8 animals, whereas HDLBP KO cells only formed tumors in 3 out of 8 mice. The tumor initiation, the tumor volume and weight was decreased upon HDLBP KO.

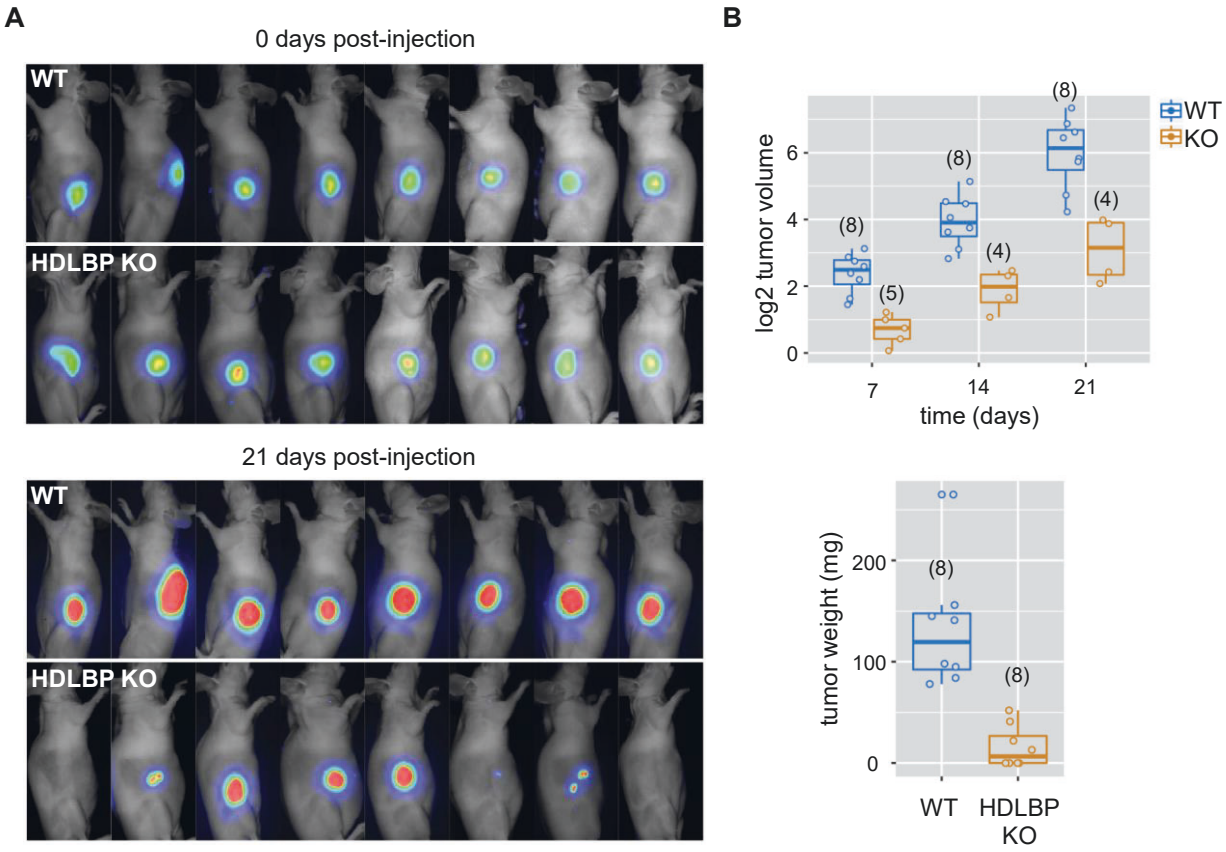

Figure 36 | HDLBP decreases tumor formation capacity *in vivo*.
(A) Non-invasive infrared RFP imaging of mice was carried out 0 days post-injection confirming that the same amount of A549 WT and HDLBP KO cells were injected (upper panel) and 21 days post-injection showing the tumors (lower panel).
(B) The tumor volume was quantified 7, 14 and 21 days post injection (upper panel) and tumor weight was quantified after 21 days post injection (lower panel). For comparison between WT and HDLBP KO an unpaired t test was used and in case of tumor absence, the weight was plotted as 0.

In conclusion, cell culture and *in vivo* experiments with HDLBP KO A549 cells revealed that HDLBP is impacting cell migration, cell proliferation and tumor formation in nude mice.

DISCUSSION

The secretory pathway is essential for proper cell functioning and key steps of the pathway are well known. This also includes the first step of the pathway where mRNAs encoding membrane or secretory proteins are targeted to the ER mainly by the SRP. It is well-established that the SRP binds ER-targeted ribosome nascent chain complexes co-translationally in the cytosol. Subsequently, a translation arrest occurs and the complex moves to the ER where the translation continues and the nascent protein is translocated into the ER lumen (Zhang and Shan, 2014). However, several mechanistic details of the SRP cycle, especially regarding the translational arrest, remain incompletely understood. Furthermore, it is discussed if the mRNA sequence itself contributes to ER-targeting. Interestingly, *cis*-acting regulatory sequence elements have been identified (e.g. Cohen-Zontag et al., 2019), but no *trans*-acting factors such as RNA-binding proteins are characterized to specifically bind these elements and to influence ER-targeting. Moreover, it is not fully understood which factors contribute to translation regulation specifically at the ER.

Based on previous studies, we hypothesized that the RNA-binding protein HDLBP is involved in these processes and characterized HDLBP by using a variety of transcriptomic and proteomic approaches and provide the most comprehensive study about the human HDLBP to date. Most importantly, we found that HDLBP influences the translation of its ER-localized target mRNAs. We estimated that HDLBP binds more than 80% of all ER-localized mRNAs by specifically recognizing a CU-containing motif and forming multivalent interactions. In addition, we discovered that ER-localized mRNAs per se differ from cytosol-localized mRNAs in their sequence composition since they contain more frequently CU-rich HDLBP binding sites. Therefore, we further defined regulatory sequence elements within ER-localized mRNAs and characterized the RNA-binding protein HDLBP, which binds to these elements, and thereby influences the translation and the efficient protein synthesis of its targets. Consequently, we simultaneously discovered a new factor contributing to translation at the ER and moreover formulate the hypothesis that HDLBP influences the SRP-mediated translation arrest. Furthermore, we showed that modulation of ER-associated translation (by HDLBP KO in A549 cells) can dramatically impact tumor formation in nude mice highlighting the broader biological significance of HDLBP.

In the following sections these findings will be discussed in more detail. First the binding characteristics of HDLBP to its mRNA targets will be addressed, next functions of HDLBP will be covered and subsequently the effect of HDLBP on tumor formation will be debated. Finally, we will speculate about how HDLBP achieves its function and will discuss mechanistic details.

HDLBP binding to mRNAs

To understand the RNA binding characteristics of the human HDLBP, we determined transcriptome-wide HDLBP binding sites using PAR-CLIP in HEK293 cells. We found that HDLBP mainly binds to the CDS of mRNAs that encode membrane and secretory proteins (Figure 5). Since it is known that these mRNAs are targeted to the ER, we investigated to which extent HDLBP binds ER-localized mRNAs. Therefore, we carried out cell fractionation experiments coupled to RNA-sequencing and defined cytosol and ER-localized mRNAs in HEK293 cells. Overlapping our PAR-CLIP identified HDLBP targets with ER-localized mRNAs allowed us to estimate that HDLBP binds more than 80% of all ER-localized mRNAs (Figure 9). Analysis of the binding motif showed that HDLBP predominantly binds to a CU-containing motif and forms high affinity multivalent interactions primarily in the CDS of ER-localized mRNAs (Figure 14, Figure 17). Therefore, ER-localized mRNAs differ from cytosol-localized mRNAs in regard to their frequency of HDLBP binding sites. These findings extent our understanding of HDLBP mRNA binding selectivity.

HDLBP binding motif

Previously, PAR-CLIP experiments using primary hepatocytes showed that the mouse homolog of HDLBP binds a CHHC or CHYC (H = A/C/U and Y = C/U) motif that is separated by 2, 5 or 8 nt (Mobin et al., 2016). Furthermore, *in vitro* assays identified that human HDLBP and its xenopus homolog bind to $(A)_nCU$ and $CU(A_)_n$ within a 75 nt long single stranded region (Kanamori et al., 1998). As a conclusion from these studies, it was suggested that HDLBP binding sites lack sequence specificity (Cheng and Jansen, 2017).

We demonstrated that HDLBP forms multivalent interactions and identified a ~40 nt long HDLBP binding site consisting of interspaced 3-4 UC/CU containing 4-mers (Figure 17). Likely, the KH domain array of HDLBP (Figure VIII) is mediating these multivalent interactions. The identified HDLBP binding site suggests that 3-4 KH domains recognize the 3-4 UC/CU containing 4-mers. However, it is not known if specific HDLBP KH domains recognize this motif or whether every HDLBP KH domain (or at least every classical KH domain) is able to bind to it. A proteome-wide screen carried out in HEK293 cells identified protein sequence regions, including HDLBP regions, that interact with mRNAs (Mullari et al., 2017). This screen is based on UV crosslinking and digestion of RNA-bound proteins followed by mass spectrometry. Thereby, HDLBP peptides corresponding to the KH domains 2, 3, 5 and 11 were recovered indicating that they contact RNA. Surprisingly, also the divergent KH domain 2 was found to bind RNA. However, further work is needed to decipher the impact and contribution of each KH domain to the HDLBP RNA substrate specificity. Interestingly, it

was reported for other KH-domain containing proteins like the IGF2BP protein family that they may recognize bipartite motifs (Biswas et al., 2019; Dagil et al., 2019). IGF2BP contains 4 KH domains and structural studies suggested that KH domain 3 recognizes ß-actin 5´-(C/A)CA(C/U)-3´ sequences, whereas KH domain 4 binds to ß-actin 5´-CGGAC-3´ sequences (Biswas et al., 2019). This highlights that KH domains can have different specificities and that future studies are necessary to fully understand the interplay of the 15 HDLBP KH domains.

Conservation of HDLBP binding specificity

The core of our identified HDLBP binding motif consists of UC/CU containing 4-mers. This is in agreement with the above mentioned reports and suggests that the binding motif of HDLBP might, to some extent, be conserved from frog and mouse to human. Moreover, also the mRNA targets of HDLBP could be conserved because the mouse homolog was shown to bind to mRNAs encoding secretory proteins (Mobin et al., 2016). However, this study suggested that HDLBP only binds to 17,5% of all mRNAs encoding secretory proteins, pointing into the direction that HDLBP has a specialized role in liver metabolism and secretory tissues. Since this statement is based on PAR-CLIP experiments using primary mouse hepatocytes, a possible explanation for this finding might be that the primary tissue limited the resolution of the experiment. In contrast, our cell culture system allowed us to use sufficient input material to obtain a very deep data set. Thereby, we identified that HDLBP is not specifically binding a small subset of mRNAs. Instead, HDLBP binds to more than 80% of all ER-localized mRNAs rather pointing to a general function for HDLBP in ER translation which could explain why HDLBP is ubiquitously expressed.

HDLBP binding motif represents a common sequence feature of ER-localized mRNAs

Furthermore, we show, in contrast to previous studies, that HDLBP mainly binds to the CDS of ER-localized mRNAs, whereas mRNAs localized in the cytosol show less HDLBP binding sites randomly distributed between the CDS and 3'UTRs (Figure 11, Figure 12). Thereby, we identified the HDLBP binding site as a common sequence feature within the CDS of ER-localized mRNAs. In doing so, we advanced the understanding of regulatory elements within ER-localized mRNAs and simultaneously described for the first time an RNA-binding protein that specifically binds to this regulatory cis-element.

Interestingly, the identified sequence feature of ER-localized mRNAs seems to overlap with the recently described SECReTE motif which consists of up to 10 nucleotide triplet repeats enriched with C and U every third base (Cohen-Zontag et al., 2019). Moreover, Cohen-Zontag et al. examined which

RNA-binding proteins can potentially bind the SECReTE motif in yeast. To this end, they investigated how many mRNA targets of a specific RNA-binding protein contain the SECReTE motif and found that ~70% of the previously identified Scp160p mRNA targets do so suggesting that the yeast homolog of HDLBP has the potential to interact with the SECReTE motif (Cohen-Zontag et al., 2019). However, no experimental evidence was provided showing that Scp160p recognizes the SECReTE motif. Cohen-Zontag and coworkers also noted that, contrary to yeast, the SECReTE motif in human transcripts preferentially occurs in UTRs especially in 3'UTRs. In contrast, our described HDLBP binding site is primarily detected in the CDS. Further analysis is needed to understand to which extent the SECReTE motif and HDLBP binding sites overlap.

HDLBP function

To determine the function of HDLBP in human cells, we pursued two hypotheses. Based on previous research, we hypothesized that HDLBP is either involved in translation regulation or in directing its targets to the ER.

Impact of HDLBP on mRNA transport to the ER

To test whether HDLBP has an effect on mRNA transport to the ER, we carried out cell fractionation experiments and separated the membrane fraction containing ER-localized mRNAs from the cytosol fraction. Subsequently, we sequenced the RNA of both fractions in HEK293 HDLBP KO and parental (WT) cells. Comparing the results between WT and HDLBP KO cells, we found no significant changes suggesting that HDLBP is not involved in mRNA localization in human cells (Figure 22). However, our global experimental approach is based on the biochemical separation of the cytosol and membrane compartment and we cannot rule out that HDLBP influences mRNA localization on a more subtle level that would not be detected with our setup. To further validate our findings and to analyze mRNA localization in more detail, one could carry out fluorescent in situ hybridization (FISH) experiments and visualize the cellular distribution of a specific HDLBP mRNA target. That would allow a high resolution, but would have the disadvantage that only limited HDLBP mRNA targets could be tested. A good candidate for FISH experiments would be the ER chaperone BiP, because we noticed for this HDLBP target less enrichment in the membrane fraction upon HDLBP KO. We tested if this is biological meaningful and carried out electron microscopy imaging and compared the ER morphology of WT and HDLBP KO cells because it was previously shown that BiP mis-regulation could lead to ER morphology defects (Li et al., 2008). However, there were no detectable ER

morphology differences between WT and HDLBP KO cells suggesting that BiP is not impaired to a level that would affect ER morphology (Figure 23).

Impact of HDLBP on translation

To test if HDLBP has an effect on translation, we carried out ribosome profiling in HEK293 WT and HEK293 HDLBP KO cells. We found that ER-localized mRNA targets of HDLBP showed less translation efficiency upon HDLBP KO suggesting that HDLBP indeed influences translation (Figure 25). To validate this finding, we performed pSILAC experiments and detected the corresponding decrease in protein synthesis of proteins encoded by mRNAs that are bound by HDLBP and ER-localized (Figure 27).

Our findings are in agreement with previous studies carried out in yeast (Baum et al., 2004; Frey et al., 2001; Hirschmann et al., 2014; Li et al., 2004; Weber et al., 1997) and mice (Mobin et al., 2016) suggesting that HDLBP is involved in translation. More precisely, Mobin and colleagues suggested that HDLBP enhances the translation of a subset of mRNAs encoding secretory proteins. This is based on their observation that several proteins showed decreased secretion from mouse primary hepatocytes upon HDLBP knockdown. However, direct evidence for the function of HDLBP as a translation enhancer was only provided for one HDLBP target by *in vitro* translation assays. In contrast to this and other studies, we show for the first time on a global scale in human cells that HDLBP influences translation and determined which mRNAs are less translated in absence of HDLBP.

To further understand how exactly HDLBP is influencing translation, we analyzed our ribosome profiling and PAR-CLIP data in more detail and made several observations.
Interestingly, we detected less ribosome profiling reads after the exposure of the signal peptide or the first transmembrane domain upon HDLBP KO (Figure 25). Thereby, we show for the first time that KO of a specific protein affects the read density in that region. According to the SRP hypothesis, one would expect that exposure of the targeting signal would cause a translation arrest and assuming that this is the case, less reads in this region can be interpreted as a reduced translation arrest caused by HDLBP KO. However, the existence of the translation arrest *in vivo* is debated in the field because it was expected that it would be detectable as a distinct peak in ribosome profiling data which was not observed in previous (Chartron et al., 2016; Schibich et al., 2016) and in our results. As explained in detail in the introduction (see “The signal recognition particle (SRP)” -> “More recent findings”) this expectation has several flaws and it is unclear at which time point exactly the translation arrest occurs and which read pattern it causes in ribosome profiling data. Therefore, we assume that the

translation arrest occurs at multiple positions along the mRNA (not causing a distinct peak) and argue that HDLBP is involved in mediating this translation arrest because we detect upon HDLBP KO significantly less reads especially in the region after exposure of the targeting signal. However, this conclusion needs further experimental validation. Therefore, it is crucial to understand on the one hand the mechanistic interaction between HDLBP and the SRP. Speculations about this interaction will be covered below (see "Unresolved questions and outlook"). On the other hand further studies are needed to clarify the specifics of the translation arrest *in vivo* because *in vitro* experiments clearly suggest that the SRP causes a translation arrest (Figure V).

Moreover, our ribosome profiling data revealed that HDLBP KO leads to higher ribosomal occupancy for various codons. We analyzed our PAR-CLIP data to understand if these codons (e.g. CUU, UUC) are also bound by HDLBP and found that this is indeed the case. Additionally, these codons are also decoded by tRNAs that are bound by HDLBP. Previously, it was shown that depletion of the yeast homolog of HDLBP decreases the relative abundance of ribosome-associated tRNAs (Hirschmann et al., 2014). To show this, Hirschmann et al. developed a protocol to detect tRNAs that co-precipitated with the ribosome and quantified the tRNA levels by real-time PCR in WT yeast and upon Scp160p depletion. As a conclusion, the authors suggested that Scp160p either increases the efficiency of tRNA recharge, or that Scp160p prevents the diffusion of discharged tRNAs. Our study supports these conclusions, but further experiments are needed to decipher the mechanistic and functional consequences of HDLBP binding to tRNAs. Nevertheless, we provide for the first time a detailed analysis of HDLBP binding to tRNAs. We determined which tRNAs are most prominently bound by HDLBP and detected that HDLBP tRNA crosslinking sites commonly occur in the variable and D loops. In addition, we found that several codons decoded by HDLBP bound tRNAs show higher ribosomal occupancy upon HDLBP KO.

Impact of HDLBP on protein secretion and cell surface expression

To monitor the impact of HDLBP on secretion in human cells, we intended to compare the secretomes of HEK293 WT and HDLBP KO cells. However, our global mass spectrometry based assessment resulted in non-quantifiable results presumably because our secretomes were contaminated by cellular proteins likely because HEK293 cells easily detach from culture plates. As a consequence, we used standard reporter assays and compared the secretion of two reporters from WT and HDLBP KO cells. Analyzing the mRNA sequences of *Gaussia* luciferase and SEAP, we identified several potential HDLBP binding sites and therefore it is very likely that HDLBP binds to mRNAs encoding the reporter

proteins. As a result, we found a reduction in *Gaussia* luciferase and SEAP secretion in HDLBP KO cells (Figure 33).

To understand the downstream consequence of HDLBP binding to mRNAs encoding membrane proteins expressed in HEK293 cells, we chose to further study one HDLBP target TFRC/CD71. We detected several crosslinking sites of HDLBP to TFRC by PAR-CLIP and also validated HDLBP binding to TFRC mRNA by real-time PCR (Figure 6). Our ribosome profiling experiments showed that the translation efficiency of TFRC was reduced and also pSILAC experiments confirmed that less TFRC protein is synthesized in the membrane fraction in HDLBP KO cells (Figure 27). In addition, FACS experiments indicated a reduction of TFRC expression on the cell surface upon HDLBP KO (Figure 34) providing evidence that HDLBP influences the surface expression of proteins encoded by mRNAs that are bound by HDLBP.

Overall, that suggests that HDLBP promotes the translation of its ER-localized mRNA targets which has functional consequences because also the protein synthesis, the secretion or surface expression are affected downstream.

HDLBP effect on tumor initiation and formation

Since HDLBP influences the secretion and surface expression of proteins encoded by mRNAs that are bound by HDLBP, we wanted to further determine the biological relevance of HDLBP. To this end, we used the lung carcinoma cell line A549 and first analyzed the viability and growth behavior of A549 HDLBP KO cells in comparison to parental A549 cells. Interestingly, HDLBP KO cells were less viable and grew slower. This encouraged us to investigate the effect of HDLBP on lung tumor initiation and formation *in vivo*. Strikingly, HDLBP KO cells only formed tumors in 3 out of 8 nude mice, whereas WT cells formed as expected tumors in all 8 mice. The remaining 3 tumors initiated by HDLBP KO cells showed a reduced tumor volume and weight suggesting that HDLBP is involved in lung tumor formation and progression. Therefore, HDLBP might be a promising therapeutic target to inhibit lung tumor growth and additional studies are needed to further explore the therapeutic potential of HDLBP and its impact on tumor biology in general.

These results are in agreement with results obtained from human hepatocellular carcinoma cells where HDLBP was shown to be required for cell proliferation and tumor growth (Yang et al., 2014). However, other studies suggest that HDLBP may function as a tumor repressor in breast cancer cell lines (Woo et al., 2011) and human sarcoma cells (Molyneux et al., 2014). In consequence of these contradictory findings, we expect that the phenotype caused by mis-regulation of HDLBP is cell type

dependent. In support of this, several disease phenotypes have been reported to be influenced by HDLBP as elaborated in the introduction (see “Disease relevance”).

Nevertheless, we speculate that most prominent phenotypes caused by dysfunctional HDLBP occur when ER translation is particularly critical. This is the case for example for specialized secretory cells in which HDLBP also shows the highest expression (Figure VII).
Furthermore, ER translation is crucial for the replication of several viruses and HDLBP KO resulted in less virus load of dengue and Zika viruses which are known to replicate in close association to the ER (Ooi et al., 2019). In contrast, the virus load of HDLBP KO cells was not altered after infection with a virus that replicates in the cytosol (Chikungunya virus) (Ooi et al., 2019; Silva and Dermody, 2017). In addition, Ooi et al. also performed infrared crosslinking and immunoprecipitation (irCLIP) of HDLBP on uninfected and virus-infected cells (dengue and Zika viruses). They found that HDLBP binds in uninfected cells mainly to mRNAs and upon virus infection more than 50% of all reads mapped to viral RNA (Ooi et al., 2019). This further illustrates the potential importance of HDLBP for virus biology and additional studies are needed to assess how this can be exploited.

Furthermore, HDLBP might influence protein folding and protein aggregation. Previously, loss of the HDLBP yeast homolog was shown to influence the aggregation of many Q/N-rich proteins and to reduce aggregation of polyQ reporters used in this study (Cheng et al., 2018). Cheng et al. speculated that this is because Scp160p slowed down the translation of the polyQ reporters. Furthermore, it was reported that modulations of the translation rate affect protein folding (Rodnina, 2016; Waudby et al., 2019) and mis-folding can cause potentially toxic protein aggregation. Here, we found that human HDLBP influences the translation efficiency of its targets and we also detected global stalling upon HDLBP KO. Therefore, we speculate that HDLBP could impact protein folding and in support of this we found key chaperones and chaperonins in close proximity to HDLBP. However, the influence of HDLBP in human cells on protein folding and possible disease implication remain to be characterized.

Unresolved questions and outlook

Possible mechanisms of how HDLBP functions in translation

HDLBP is a cytoplasmic protein in HEK293 cells that co-localizes with the ER where it influences the translation of its mRNA targets. These mRNA targets are mainly bound by HDLBP in their CDS. Moreover, our data also provide evidence that HDLBP binds to tRNAs, ribosomes, the SRP and is primarily in close proximity to proteins associated with translation. How do these interactions connect up to create a conclusive hypothesis on how HDLBP functions in translational control?

After the transcription of an mRNA, the mRNA is processed and exported from the nucleus into the cytosol. Here, the translation initiates and the ribosome-nascent-chain complex either first moves to the ER or directly continues the translation in the cytosol. Due to the cellular localization of HDLBP and the fact that it prefers to bind to the CDS, the time window in which HDLBP influences its mRNA targets is rather short. On one hand, it can only happen after the mRNA is exported from the nucleus because we detect HDLBP only in the cytoplasm. On the other hand, HDLBP binding to the CDS of its targets is most likely interrupted once the ribosome starts translating the mRNA. Therefore, we hypothesize that HDLBP is predominantly important for the pioneer round of translation after which it is removed.

Crucial for the pioneer round of translation is that, upon nuclear export, several complexes are still bound to the mRNA (Maquat et al., 2010; Woodward et al., 2017). This includes the exon junction complex, which for example, ensures proper functionality of each mRNA because nonfunctional mRNAs will be identified and subsequently eliminated. Moreover, exported mRNAs are bound by the nuclear cap-binding protein heterodimer CBP80-CBP20. This complex is needed for the loading of ribosomes on the mRNA and is afterwards replaced by eIF4E (Maquat et al., 2010). We do not find evidence that HDLBP is involved in these first steps of the pioneer round of translation because no important factors like the exon junction complex are in close proximity to HDLBP as determined by our BioID experiment. However, several translation initiation factors including eIF4E are in our BioID data set suggesting that HDLBP gets involved once eIF4E directs translation initiation. Next, the majority of mRNAs bound by HDLBP are targeted to the ER where the final protein is synthesized. Therefore, it is likely that HDLBP functions between translation initiation, ER-targeting and the first round of translation. This also suggests that HDLBP modulates active translation. In a proteome-wide screen in MCF7 cells, HDLBP was identified to show significantly decreased RNA-binding upon arsenite-induced global translation arrest which supports this notion (Trendel et al., 2019). However,

we only have limited data to further speculate about mechanistic details about the role of HDLBP in translation and several scenarios are conceivable.

Is HDLBP influencing translation initiation and/or elongation?

It could be possible that HDLBP directly influences translation initiation and/or elongation. Data supporting this are mainly our ribosome profiling identified global stalling observations upon HDLBP KO. We detected higher read coverage at P sites but most prominently at E sites. This could be an argument that HDLBP is influencing translation elongation, because it was suggested that the E site functions as a sensor of ribosome elongation kinetics (Schuller and Green, 2018). Therefore, global stalling mostly at E sites upon HDLBP KO might be consequence of a modified translation elongation rate during the pioneer round of translation. Moreover, a role of HDLBP in translation elongation would be in agreement with its mRNA binding preference (all over the CDS). However, we only detected HDLBP in close proximity to 2 translation elongation factors (eEF1B2, eEF1D). In contrast, we found 11 different translation initiation factors in close proximity to HDLBP. Therefore, HDLBP could also enhance translation initiation by simultaneously binding to its mRNA targets and to ribosomes leading to a higher ribosome load on HDLBP mRNA targets. Further experiments are needed to decipher how HDLBP influences translation initiation and/or elongation.

Is HDLBP influencing the SRP cycle?

Furthermore, it could be possible that HDLBP influences the ER-targeting step and thereby promotes the translation of its ER-localized targets. Most mRNAs are targeted to the ER by the SRP. However, also mRNAs with no known targeting signal are localized to the ER probably in a SRP-independent manner. Interestingly, HDLBP contributes to efficient translation of both SRP-dependent and –independent mRNAs. Since it is not known how exactly SRP-independent mRNAs are targeted to the ER it is hard to speculate how HDLBP contributes to their efficient translation. In contrast, speculations about the contribution of HDLBP to the SRP cycle are more profound. From structural considerations, it would be conceivable that HDLBP stabilizes the SRP-ribosome-mRNA interaction and thereby also contributes to translation arrest. Data that support this hypothesis are mainly our PAR-CLIP identified HDLBP crosslinking sites detected in the 7SL RNA and the 40S ribosomal subunit which are in proximity to translation initiation and elongation factor binding sites. Therefore, HDLBP binding to these sites could influence the translation arrest. Besides these structural considerations, we also found reduced ribosomal occupancy around targeting signals upon HDLBP KO moreover suggesting that HDLBP influences translation arrest. However, this hypothesis would not explain

why HDLBP binds its ER-localized mRNA targets all over the CDS and not predominantly around the first targeting signal. In addition, one would expect that a non HDLBP stabilized SRP-ribosome interaction would not only impair efficient translation, but furthermore would result in less efficient ER-targeting of the complex which is not supported by our data.

To further decipher the contribution of HDLBP to the SRP cycle it is fundamental do understand when exactly HDLBP affects its mRNA targets in relation to the SRP. Does it affect its targets during translation initiation before the SRP is recruited? Could HDLBP be a "marker" for ER-destined mRNAs and could thereby help the SRP to bind to the correct ribosome-nascent-chain complexes? Is the HDLBP-SRP interaction crucial for the function of HDLBP or does HDLBP influence its targets after the SRP dissociates by modulating the translation elongation rate?
First step, in order to answer these questions, would be to determine ribosome-nascent chain complexes that are bound by the SRP in a human system. Since selective ribosome profiling of SRP-bound ribosomes was already successfully performed in yeast (Chartron et al., 2016) it would be insightful to establish this experiment in human cells. Next, the same experiments could be carried out using HDLBP KO cells. Comparing both data sets would help to identify if HDLBP KO results in a different SRP binding pattern which would help to further characterize the HDLBP-SRP interaction.

Overall, we demonstrate that the highly conserved and ubiquitously expressed protein HDLBP has a general function in the translation of ER-localized mRNAs and future research will clarify the mechanistic details and exploit the therapeutic potential of HDLBP.

MATERIALS AND METHODS

Related to results chapter [1]

Cell lines and culture conditions

HEK293 Flp-In T-Rex cells (HEK293) (Thermo Fisher Scientific), HEK293 stable cell lines, HDLBP KO cell lines, A549 cells and A549 HDLBP KO cell lines were grown in standard Dulbeco´s modified Eagle´s medium (DMEM) (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 1% L-glutamine (200 mM, Thermo Fisher Scientific).

Stable HEK293 cell lines expressing BirA-FLAG-HDLBP or FLAG/HA-HDLBP were generated by hygromycin selection (Gregersen et al., 2014). Induction of the stable cell lines was achieved by adding 1 µg/ml of doxycycline (DOX) to the culture medium for ~16 h.

HEK293 and A549 HDLBP knockout cell lines were generated using the Edit-R CRISPR-Cas9 Gene Engineering kit (Dharmacon) according to manufacturer´s protocol. Briefly, transfections of synthetic tracr RNA (U-002000-05), hCMV-PuroR-Cas9 (U-005100-120) and predesigned HDLBP crRNA (either guide 1 (CR-019956-01-0005, target sequence: TGCTCCTCCTGCAGATTAAC) or guide 2 (CR-019956-04-0005, target sequence: GAGCGACCCTCCAACCTACA)) were performed using DharmaFECT Duo transfection reagent (T2010-01) in a 12-well plate. Cells were transferred after 2 days to a 10 cm dish and treated with puromycin (2 µg/ml for HEK293 cells and 1 µg/ml for A459 cells). The surviving colonies were selected and analyzed by Western Blot.

Western Blot

Cell pellets were resuspended directly in Laemmli buffer, sonicated (5-second-pulse at 80% amplitude) and centrifuged (10000 g, 10 min, 4°C). Before resolving the proteins by SDS-PAGE the samples were boiled for 3 min (95°C). To transfer the proteins to a nitrocellulose membrane (Whatman) a semi-dry blotting apparatus (20 V for 1 h) was used. The membranes were blocked with 5% non-fat milk for 1 h and subsequently incubated with the primary antibody overnight. After washing the membranes 3 times with TBST (150 mM NaCl, 20 mM Tris-HCl [pH 7.5], 0,1% Tween20) for 5 min the membranes were incubated with horseradish peroxidase conjugated secondary antibody for 2 h. The membranes were washed 3 times with TBST for 5 min, bands were visualized with an ECL detection reagent (GE Healthcare) and imaged with a LAS-4000 imaging system or an Amersham Imager 680 imaging system (GE Healthcare).

Related to results chapter [2]

Immunohistochemistry

HEK293 cells were grown on coverslips in 6-well plates. ER-Tracker Red (Thermo Fisher Scientific) was used according to manufacturer´s instructions. The medium was removed and cells were washed with PBS. Cells were fixed with 3,7% formaldehyde in PBS for 10 min at room temperature. After washing the cells twice with PBS, cells were permeabilized with 0,2% Triton X-100 in PBS for 10 min at room temperature. Cells were washed twice with 0,05% Tween20 in PBS (PBST) and blocked with 10% FBS in PBST for 30 min at room temperature. The primary antibody (HDLBP) was diluted 1:500 in PBST and incubated over night at 4°C. After washing twice with PBST, the secondary antibody was applied (Alexa Fluor 488 anti-rabbit, 1:2000 in PBST) for 1 h at room temperature. Cells were washed 5 times with PBST. Hoechst staining (1:2000 in PBST, stock concentration: 1mg/ml) was carried out for 15 min at room temperature. After washing twice with PBS, cover slips were mounted with ProLong Gold Antifade Mountant (Thermo Fisher Scientific) and imaged with a Keyence microscope (BZ-X700).

Related to results chapter [2-6]

Cell fractionation

Cell fractionation by sequential detergent extraction was carried out as previously described (Jagannathan et al., 2011) with minor modifications. HEK293 and HEK293 HDLBP knockout cell lines (1 dish (15 cm) per replicate) were grown to ~90% confluency and washed with PBS. All further steps were performed on ice using ice cold reagents and cells were always pelleted at 3000 g for 5 min at 4°C. First, PBS containing 50 µg/ml cycloheximide was added for 10 min and in the meantime cells were scraped. After pelleting, the cells were resuspended with 500 µl permeabilization buffer (110 mM KOAc, 25 mM K-HEPES [pH 7. 2], 2.5 mM $Mg(OAc)_2$, 1 mM EGTA, 0,015% digitonin, 1 mM DTT, 50 µg/ml cycloheximide, complete EDTA-free protease inhibitor cocktail [Roche], 40 U/mL SUPERaseIn (Thermo Fisher Scientific)) per sample. After 15 min of incubation at 4°C on a rotating wheel, the cells were pelleted. The supernatant was collected as the cytosolic fraction. The pellet was washed with 5 ml of washing buffer (110 mM KOAc, 25 mM K-HEPES [pH 7.2], 2.5 mM $Mg(OAc)_2$, 1mM EGTA, 0,004% digitonin, 1 mM DTT, 50 µg/ml cycloheximide). After pelleting, cells were resuspended with 500 µl lysis buffer (400 mM KOAc, 25 mM K-HEPES [pH 7.2], 15 mM $Mg(OAc)_2$, 0,5% IGEPAL CA-630, 1 mM DTT, 50 µg/ml cycloheximide, complete EDTA-free protease inhibitor cocktail, 40 U/mL SUPERaseIn) per sample. After 5 min of incubation, cells were pelleted and the supernatant was collected as the membrane fraction. Next, fractions were clarified by centrifugation

at 7500 g for 10 min at 4°C. 250 µl of each sample was collected and subsequently the RNA was isolated using Trizol LS (Thermo Fisher Scientific) in combination with RNA Clean & Concentrator-25 kit (Zymo Research) for RNA-sequencing (see "RNA-sequencing library preparation").

RNA-sequencing and data analysis

The TruSeq Stranded mRNA kit (Ilumina) was used according to manufacturer´s protocol with 1 µg of RNA as input for preparation of RNA sequencing libraries. Next generation sequencing runs were performed on a HiSeq 4000 or NextSeq 500 Illumina instruments. After demultiplexing the reads, the 3'adapter sequences were removed by Flexbar (v2.5). Read counts per gene and TPM values were calculated by RSEM (v.1.2.20) (Li and Dewey, 2011) using default parameters and Bowtie (v.1.1.2) (Langmead et al., 2009) as transcriptome alignment program. For normalization raw read counts and pairwise comparisons between and within fractions DESeq2 (Love et al., 2014) was used with standard parameters. The DESeq2 output provided log2-transformed fold changes. Membrane-to-cytosol enrichment was equivalent to the log2-transformed fold change between membrane and cytosolic samples. To classify membrane-localized and cytosolic mRNAs, we used cutoffs of >=1.5 and <=0.5, respectively, for all mRNAs with TPM>=10.

Targeting signal annotations

Signal peptide and transmembrane helix annotations were obtained from Ensembl Biomart (http://grch37.ensembl.org/biomart/martview), which uses SignalP (Bendtsen et al., 2004) and TMHMM (Krogh et al., 2001) for annotation. Annotations were downloaded in protein sequence coordinates which were converted into transcript coordinates by Biostrings (v2.52.0, Bioconductor). Based on these annotations, we classified tail-anchored proteins as those proteins lacking signal peptide, for which the first transmembrane helix started 50 or less amino acids from the C-terminus. We classified mitochondrial DNA encoded proteins as those that contained an "MT-" prefix in their gene symbol. A list of mitochondrial proteins encoded in the nuclear DNA was downloaded from Mitocarta (Calvo et al., 2016).

PAR-CLIP and data analysis

Experimental conditions

PAR-CLIP was carried out as described previously (Hafner et al., 2010) with minor modifications. Stable cell lines expressing FLAG/HA-HDLBP were grown with 100 µM 4sU for 16 h (20 dishes (15 cm) per replicate). Cells were UV crosslinked (365 nm, 0.15 J/cm^2), harvested, snap frozen and

subsequently stored at -80°C. Cell pellets were incubated for 30 min on ice with 3 cell pellet volumes of lysis buffer (50 mM Tris-HCl [pH 7.4], 100 mM NaCl, 1% IGEPAL CA-630, 0,1% SDS, 0,5% sodium deoxycholate, complete EDTA-free protease inhibitor cocktail [Roche]). Cell lysates were cleared (13000 rpm, 15 min, 4°C) and filtered. RNase T1 was added at final concentration of 1 U/µl for 10 min at 22°C. 20 µl beads (Dynabeads Protein G, Thermo Fisher Scientific) per 1 ml lysis buffer were washed twice with lysis buffer and incubated for 1 h with 0,25 µg/µl FLAG antibody (Sigma-Aldrich, F1804). Afterwards beads were washed twice with lysis buffer and added to the sample. After 2 h incubation at 4°C on a rotating wheel the beads were washed with IP wash buffer (50 mM HEPES-KOH [pH 7.5], 300 mM KCl, 0,05% IGEPAL CA-630, 0,5 mM DTT, complete EDTA-free protease inhibitor cocktail) 3 times. A second RNase T1 treatment was carried out and for replicate 1, a final concentration of 1 U/µl and for replicate 2 a final concentration of 15 U/µl was used for 10 min at 22°C. Subsequently, beads were washed with high-salt wash buffer (50 mM Tris-HCl [pH 7.4], 1 M NaCl, 1% IGEPAL CA-630, 0,1% SDS, 0,5% sodium deoxycholate, 1 mM EDTA, complete EDTA-free protease inhibitor cocktail [Roche]) 3 times. Afterwards beads were washed 2 times with polynucleotide kinase buffer (50 mM Tris-HCl [pH 7.5], 50 mM NaCl, 10 mM $MgCl_2$, 5 mM DTT). Next, samples were radiolabeled using T4 polynucleotide kinase (NEB) and resolved on a Novex 4-12% BisTris gel (Thermo Fischer Scientific). The HDLBP-RNA complexes were transferred to a nitrocellulose membrane (Whatman), exposed for 30 min to a phosphorimager screen and excised at approximately 160 kDa. After proteinase K (Roche, 40 min at 50°C) digestion the RNA was purified by phenol-chloroform treatment and precipitation. For preparation of sequencing libraries the RNA was firstly ligated to a 3' adapter 4N-RA3 (see "Oligonucleotides") and gel-purified using a 15% denaturing urea-PAGE gel (Carl Roth). Next, the 5' adapter OR5-4N (see "Oligonucleotides") was ligated and gel purified using a 17% denaturing urea-PAGE gel. The RNA was reverse transcribed and PCR-amplified using Phusion High-Fidelity DNA polymerase (Thermo Fischer Scientific) (see "Oligonucleotides"). The cDNA was visualized on a 2,5% agarose gel and a 140-160 bp sized fragment was excised which was purified using the Zymoclean Gel DNA Recovery kit (Zymo Research). Next generation sequencing was performed on a HiSeq 2500 Illumina instrument (1x51+7 cycle).

Data analysis

Reads were demultiplexed and the 3'adapter sequence was removed by Flexbar (v2.5). Next, reads were collapsed to remove PCR duplicates. This was followed by trimming of 4 nucleotides from both 5' and 3' end of the read by FASTX Toolkit v0.0.14. The reads were aligned to the human genome (hg19 build) using BWA v0.7.15-r1140 and the computational PAR-CLIP pipeline (v0.97a) (https://github.com/marvin-jens/clip_analysis) (Lebedeva et al., 2011). Briefly, read clusters were called from unique alignments and scored for characteristic T-to-C conversions. After false positive

filtering (using antisense clusters as a decoy database and a false discovery rate of 0.05) the remaining clusters were kept and written as bed files. In addition, clusters obtained from each biological replicate were filtered for reproducibility and only those clusters were considered that overlapped by at least 50% of their nucleotide length between replicates. Furthermore, we required that the positions of the highest T-to-C conversion values per cluster were no more than 10 nt apart between 2 replicates. Each kept cluster was also required to have at least 3 or more mean T-to-C conversions calculated between replicates. To obtain gene level binding information, we summed T-to-C conversions in reproducible clusters for each gene. Thereby, we obtained the total number of crosslink positions within whole mRNAs, or within the CDS, 5'UTR or 3'UTR. To correct for PAR-CLIP expression level bias, we divided total number of crosslinks per gene by the corresponding TPM value which was obtained by RSEM (v.1.2.2) from our RNA-sequencing experiment from total HEK293 cells (see "RNA-sequencing and data analysis" description). As a result we obtained our PAR-CLIP enrichment values.

To assess the relative distribution of T-to-C conversions within the transcriptome and to obtain accurate mapping to transcripts originating from repetitive genomic loci, reads were sequentially mapped to reference transcripts by Bowtie2 (v2.3.2). This was done in the following order by retaining the unmapped reads from the previous to the next mapping step. First, mapping to human pre-rRNA (GenBank U13369.1) was carried out, followed by rRNA (GenBank NR_023363.1, NR_003285.2, NR_003287.2, NR_003286.2), snRNA, snoRNA, other ncRNA (all from Ensembl, including RN7SL), tRNA (GtRNADb), mtDNA (GenBank AF347015.1) and finally the human genome (hg19, iGenomes). The last genome-mapping step was carried out using the STAR aligner (v2.2.1) where only uniquely mapped reads (MAPQ=255) were retained. Except for the tRNA mapping (see "Mapping to reference tRNAs and quantification" description), we retained reads that mapped with a Bowtie2 MAPQ=20 or more and T-to-C conversions were extracted using row_mpile_coverage_plus_TC.pl script (Schueler et al., 2014).

For transcriptome level analysis, the remaining reads (after mtDNA mapping) were aligned to the transcriptome sequence (GTF annotation file Gencode v19) using the STAR aligner (v2.2.1) by retaining only the reads that uniquely mapped to the hg19 genome (parameters --outFilterMultimapNmax 1 --outFilterMismatchNmax 5 --outFilterMatchNmin 15 --alignSJoverhangMin 5 --seedSearchStartLmax 20 --outSJfilterOverhangMin 30 8 8 8 --quantMode TranscriptomeSAM). For further analysis we considered only 1 transcript isoform per gene. This filtering was carried out based on RSEM (v1.2.2) results derived from our RNA-sequencing experiment from total HEK293 cells (see "RNA-sequencing and data analysis" description). Thereby, only the most highly expressed transcript isoform for a given gene was retained. T-to-C conversions

with transcriptome coordinates were obtained from the BAM file using the row_mpile_coverage_plus_TC.pl script (Schueler et al., 2014) and the output was written into a bed file. BedTools was used to intersect and to only keep reproducible T-to-C positions that were present in both replicates. Among those, we excluded T-to-C positions that were highly likely to be point mutations by retaining only those that had a transition specificity (positional T-to-C conversion count vs. total read coverage) lower than 0.95. To assess the positional HDLBP crosslinks across 5'UTR, CDS and 3'UTR, we used reproducible T-to-C conversions and normalized them for library size. Thereby, we obtained T-to-C conversions per million. Only transcripts with a total transcript T-to-C conversions per million of at least 5 in both replicates were included. For each T-to-C position we next derived the relative positions within 5'UTR, CDS or 3'UTR and divided the T-to-C per million value with the maximum T-to-C per million value of the corresponding transcript. For each position, we averaged the scaled T-to-C per million values over all transcripts and plotted them in regard to their nucleotide positions within 5'UTR, CDS or 3'UTR. Classification of membrane- and cytosol-localized mRNAs was carried out according to the results of our RNA-sequencing cell fractionation experiment.

K-mer enrichment and multivalency analysis

For these analyses, we made use of the reproducible T-to-C conversions per million values from the transcriptome level analysis as described in the PAR-CLIP data analysis section. To calculate the frequency of crosslinked k-mers (4-12 nt in length), we counted all possible k-mers at any crosslinking position for all membrane and cytosol-localized transcripts in their 5'UTRs, CDS and 3'UTRs. The frequency was obtained by dividing with the total detected k-mer number in the whole sample. Next, k-mers were then ranked by their total frequency in the transcriptome. Top 10 k-mer frequencies were displayed according to mRNA localization and transcript region (5'UTR, CDS, 3'UTR).

For each transcript and its regions, we obtained k-mer specific crosslinking quantification by summing up the T-to-C conversion counts for each possible k-mer within a CDS and 3'UTR. In order to normalize this metric, we divided the values with the length of the respective CDS/3'UTR and with the transcript expression level as quantified by the total RNA-sequencing experiment (see "RNA-sequencing and data analysis" description). Next, the k-mers were ranked according to the median of the log2-transformed normalized crosslinked k-mer signal.

Sequence analyses of differentially localized transcripts and their regions were performed by calculating the k-mer frequency using the oligonucleotideFrequency function in Biostrings (v2.52.0, Bioconductor). We used fasta sequences of the same transcript subset as selected for PAR-CLIP and

ribosome profiling analyses. Thereby, we only retained the most highly expressed isoform per gene. For every k-mer length, we calculated the difference in frequency obtained from transcripts with differing localization and/or region (e.g. difference between cytosolic CDS and membrane CDS frequencies). Next, we computed the z-scores of the differences in frequency for the top 40 HDLBP crosslinked k-mers (as described above) and all other k-mers to compare between different k-mer lengths. Wilcoxon rank sum tests were used pairwise to compare z-score distributions between different k-mer lengths.

To obtain multivalency scores for crosslinked 4-mers, we counted all crosslinked 4-mers in the region +40/-40 nt around each T-to-C conversion position by using the findOverlaps function of the Genomic Ranges package (v.1.36.0, Bioconductor) to obtain the nucleotide distances between all T-to-C conversions within a transcript. The closest region (+4nt/-4nt) around the reference T-to-C conversion was excluded from counting, because it would otherwise dominate the total crosslink signal within the +40/-40 nt regions. Next, 4-mers were ranked by the frequency of detected 4-mers, which was obtained by dividing with the total number of all detected 4-mers within the +40/-40 nt regions excluding the closest region. For the top10 enriched 4-mers by frequency, we then binned the multivalency scores into 5 equally-sized categories and compared the total normalized T-to-C conversion signal over the +40/-40 nt regions. Next, we calculated the percentage of total T-to-C conversions for every nucleotide position within the +40/-40 nt region for each multivalency bin.

Finally, we determined the multivalency potential of sequences of differentially localized transcripts and their regions. We chose two 4-mer groups, namely a positive set (UUCU), which consisted of top 10 crosslinked 4-mers, and a negative set (AAGU). The negative set had a similar transcriptome frequency as the positive set but was not enriched in HDLBP PAR-CLIP. Next, we counted the occurrence of these 2 sets in 30-nt sliding windows in transcript sequences by using vcountPDict function in Biostrings. We kept all windows with at least three 4-mer group counts and summed them over transcript regions to obtain their frequency. Next, the distribution of the frequencies was compared between differentially localized transcripts and their regions with Wilcoxon rank sum tests.

RNA co-immunoprecipitation and real-time PCR

HEK293 stable cell lines expressing FLAG/HA-HDLBP were scraped, centrifuged in a capped syringe and frozen by pressing the cells through the syringe directly in liquid nitrogen. Frozen cells were grinded, aliquoted as powder and subsequently stored at -80°C. 200 mg cryopowder were resuspended with 1 ml of lysis buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 0,5% IGEPAL CA-630, complete EDTA-free protease inhibitor cocktail [Roche]) and incubated for 30 min on ice. In the

meantime samples were sonicated (5-second-pulse at 20% amplitude). Cell lysates were cleared (13000 rpm, 15 min, 4°C) and 25 µl beads (Dynabeads Protein G, Thermo Fisher Scientific) per 1 ml lysis buffer were washed twice with lysis buffer and coated for 1 h with 0,25 µg/µl FLAG antibody (Sigma-Aldrich, F1804) or with 0,25 µg/µl IgG antibody (Sigma-Aldrich, M5284). Next, beads were washed with lysis buffer twice and added to the sample. After 2 h incubation on a rotating wheel at 4°C the beads were washed 5 times with lysis buffer and Laemmli buffer was added for Western Blot analysis or Trizol LS (Thermo Fisher Scientific) for RNA isolation. After RNA isolation a DNase (NEB, M0303) treatment and reverse transcription using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific, 18080093) was performed according to manufacturer's protocol. Real-time PCR was carried out using SYBR Green PCR Master Mix (Thermo Fisher Scientific, 4309155) and primers listed under "Oligonucleotides". Finally, the fold enrichment was calculated from Ct values detected in anti-FLAG and IgG control samples (2^(anti-Flag CT value – IgG control CT value)).

Related to results chapter [6]

Mapping PAR-CLIP crosslinks to ribosome and SRP 3D structures

For this analysis, we made use of our pre-rRNA and RN7SL alignments and extracted T-to-C conversions as described above in the PAR-CLIP section. Next, we aligned the pre-rRNA and RN7SL fasta files to the 18S rRNA and RN7SL fasta files which we obtained from the published structures (PDB id's 4V6X, 6FEC, 3JAJ) (Anger et al., 2013; Eliseev et al., 2018; Voorhees and Hegde, 2015) and defined the orthologous position of the T-to-C conversion. We used Pymol (v.2.3.3) to simultaneously visualize translation initiation factors and expansion segments, as well as the SRP and expansion segments. Therefore, we juxtaposed 4V6X and 6FEC structures and in addition the 4V6X and 3JAJ structures using the align command. We labeled crosslinked nucleotides and other structural features accordingly.

Additional published datasets

IRE1 PAR-CLIP (Acosta-Alvear et al., 2018) was obtained by communication with the corresponding author and SSB PAR-CLIP datasets (Garzia et al., 2017; Gogakos et al., 2017) were obtained from GEO under accession GSE95683 and from SRA under accession SRR4301753. MOV10 PAR-CLIP (Gregersen et al., 2014) was downloaded from GEO using the accession GSE48245.

BioID and data analysis

Experimental conditions

The BioID proximity ligation assay was carried out as described before (Couzens et al., 2013) with minor modifications. HEK293 stable cell lines expressing BirA-FLAG-HDLBP (4 dishes (15 cm) per replicate) were incubated in the absence or presence of 1 µg/ml doxycycline for 24 h. Next, 250 µM biotin was added for 3.5 h. The cells were washed 4 times with PBS, harvested, snap frozen in liquid nitrogen and subsequently stored at -80°C. Cell pellets were incubated with 3 ml RIPA buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1% IGEPAL CA-630, 1 mM EDTA, 1 mM EGTA, 0,1% SDS, 0,5% sodium deoxycholate, complete EDTA-free protease inhibitor cocktail [Roche]) per sample for 10 min on ice. Next, the cell lysates were passed 8 times through a 21 G needle and sonicated (6 times 5-second-pulses at 30% amplitude). 250 U benzonase (Merck Millipore) was added per sample for 1 h on ice at slow agitation. The cell lysates were cleared (15000 g, 15 min, 4°C) and filtered through a 5 µm Supor membrane. To wash streptavidin sepharose (GE Healthcare, 17-5113-01), the suspension was centrifuged at 400 g for 1 min, the supernatant was removed and the sepharose was resuspended with 1 ml RIPA buffer (without 0,5% sodium deoxycholate and without complete EDTA-free protease inhibitor cocktail). This wash step was repeated 2 times. Next, the RIPA buffer was removed and 40 µl sepharose was added was added to the sample. After 3 h incubation on a rotating wheel at 4°C the sepharose was washed twice with RIPA buffer, twice with TAP lysis buffer (50 mM HEPES-KOH [pH 8.0], 100 mM KCl, 10% glycerol, 2 mM EDTA, 0,1% IGEPAL CA-630) and 3 times with 50 mM ammonium bicarbonate. 90% of each sample was stored at -80°C and further processed for mass spectrometry. An aliquot (10% of the sample) was used for Western Blot analysis for quality control. Therefore, we incubated the sample at 90°C for 5 min in 95% formamide containing 10 mM EDTA and 1 mM biotin and subsequently Laemmli buffer was added.

Mass spectrometry

The beads were resuspended in 200 µl of 50 mM ammonium bicarbonate containing 2 µl trypsin (Promega, V511A). Samples were incubated for 16 h in a Thermomixer at 37°C and 750 rpm shaking. Afterwards, 1 µg of trypsin was added and the samples were incubated further for 2 h. Next, samples were centrifuged at 400 g for 2 min and the supernatant transferred to a new vial. To ensure complete bead removal, the samples were centrifuged at 16000 g for 10 min. The supernatant was transferred to a new vial containing 5 µl of trifluoroacetic acid. Each sample was loaded on two StageTips (Rappsilber et al., 2003) for desalting. Eluates of each sample were pooled together prior to MS analysis.

For all the samples, 5 µl were injected in duplicates on a LC-MS/MS system (EkspertNanoLC 415 [Eksigent] coupled to Q Exactive HF [Thermo]), using a 240 min gradient ranging from 2% to 45% of

solvent B (80% acetonitrile, 0,1% formic acid; solvent A= 5% acetonitrile, 0,1% formic acid). For the chromatographic separation a 30 cm long capillary (75 um inner diameter) was packed with 1.8 micron C18 beads (Reprosil-AQ, Dr. Maisch). On one end of the capillary nanospray tip was generated using a laser puller (P-2000 Laser Based Micropipette Puller, Sutter Instruments), which allowed fretless packing. The nanospray source was operated with a spay voltage of 2.3 kV and an ion transfer tube temperature of 260°C. Data were acquired in data dependent mode using the top10 method (one survey MS scan with resolution 60000 at m/z 200, followed by up to 10 MS/MS scans on the most intense ions, resolution 15000, intensity threshold 5000). Once selected for fragmentation, the ions were excluded from further selection for 30 sec to increase new sequencing events.

The raw data were analyzed using the MaxQuant proteomics pipeline (v1.5.3.30 and v1.5.8.3) and the built in the Andromeda search engine (Cox et al., 2011) with the Uniprot Human database. Carbamidomethylation of cysteines was chosen as fixed modification, oxidation of methionine and acetylation of N-terminus were chosen as variable modifications. The search engine peptide assignments were filtered at 1% FDR and the feature match between runs was not enabled. Other parameters were left as default.

Mass spectrometry proteomics data have been uploaded to ProteomeXchangeConsortium (http://proteomecentral.proteomexchange.org) via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD018313.

Data analysis

First, we excluded those protein groups that were identified by MaxQuant as potential contaminants and by reverse peptide sequences. We requested that the number of razor unique peptides per protein group was at least 3. The fold change enrichment in HDLBP BioID was calculated by dividing the LFQ values from the doxycycline (Dox)-treated condition with untreated condition-derived LFQ values. We obtained 3 biological replicates (batches 1, 2 and 3) of the Dox-treated cells and only 2 biological replicates of the untreated controls (batches 2 and 3). Therefore, we divided the Dox batch 1 with untreated batch 2 to finally obtain enrichment values for 3 replicates. In order to be defined as enriched in the HDLBP BioID, we requested that the protein enrichment value was at least 3 and that the log2-transformed LFQ intensity in the Dox-treated condition was at least 27. For proteins where the calculation of enrichment values was not possible due to the fact that LFQ intensity in the untreated condition equaled zero, we requested that the log2-transformed LFQ intensity in the Dox-treated condition was at least 27. These filtering criteria were applied separately for each biological replicate and therefore the protein group had to fulfill them in all 3 replicates to be defined as enriched. Next, we ranked the enriched proteins by averaging their LFQ intensity in the Dox condition over 3 replicates. To visualize the top 60 enriched protein groups by mean LFQ, we additionally

displayed the mean log2-transformed enrichment values. For those proteins that did not have an enrichment value due to the fact that LFQ intensity in the untreated condition equaled zero, we used a maximum of the mean enrichment value and added a value of 2.

Related to results chapter [7]

Electron microscopy

HEK293 WT and HDLBP knockout cell lines (1 (15 cm) dish per replicate) were grown to 70% confluency. Cells were fixed with fixation buffer (2% formaldehyde, 1% glutaraldehyde in 0,1 M phosphate buffer) for 24 h. After treating the cells with 1% osmium tetroxide for 3 h they were dehydrated through a graded series of ethanols and embedded in Poly/Bed 812 (Polysciences). Ultrathin sections were stained with uranyl acetate and lead citrate and examined with a Morgagni electron microscope (Thermo Fisher Scientific). Digital images were taken with a Morada CCD camera and the iTEM software (EMSIS).

Related to results chapter [8-9]

Ribosome profiling and data analysis

Experimental conditions

Ribosome profiling was carried out as described previously (Ingolia et al., 2009) with minor modifications. HEK293 WT and HDLBP knockout cell lines (1 10 cm dish per replicate) were grown to ~90% confluency. The cells were washed with ice cold PBS containing 100 µg/ml cycloheximide. The PBS was thoroughly removed. Plates were put on liquid nitrogen for 10 sec and subsequently on ice. 400 µl mammalian polysome buffer (20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/ml cycloheximide, 1% Triton X-100 and 25 U/mL TurboDNase [Thermo Fisher Scientific]) per plate was added. The cells were scraped, collected and the lysates were passed 10 times through a 26 G needle. After clearing the cell lysates by centrifugation (20000 g, 10 min, 4°C) 120 µl cell lysate aliquots were snap frozen and subsequently stored at -80°C. One aliquot of cell lysate was used for RNA-sequencing (see "RNA-sequencing library preparation"). Therefore, the RNA was isolated using Trizol LS (Thermo Fisher Scientific) in combination with RNA Clean & Concentrator-25 kit (Zymo Research). Another aliquot of cell lysate was used to isolate the ribosome-protected fragments by adding 300 U RNase I (Thermo Fisher Scientific) for 45 min at slow agitation at room temperature. Meanwhile, MicroSpin S-400 HR Columns (GE Healthcare) were equilibrated by adding regularly ice cold mammalian polysome buffer (without DTT, cycloheximide, Triton X-100, TurboDNase) to the columns. Next, the columns were centrifuged (600 g, 4 min, 4°C). 100 U

SUPERaseIn (Thermo Fisher Scientific) per sample was added, mixed and the cell lysates were applied dropwise to the columns (100 µl cell lysate per column). The columns were centrifuged (600 g, 2 min, 4°C). The flow-through was collected and the RNA was isolated using Trizol LS (Thermo Fisher Scientific) in combination with RNA Clean & Concentrator-25 kit (Zymo Research). Next, the ribosome-protected fragments were depleted with the RiboZero Kit (Illumina) according to manufacturer´s protocol using 5 µg RNA as input. The remaining RNA was separated on a 17% denaturing urea-PAGE gel (Carl Roth). RNA fragments in the range from 27 nt to 30 nt were excised (defined by markers (see "Oligonucleotides")). Sequencing libraries were generated as described in the PAR-CLIP section and next generation sequencing was performed on a HiSeq 2500 Illumina instrument (1x51+7 cycle).

Data analysis

Reads were demultiplexed and adapter sequences were removed by Flexbar (2.5). Next, the reads were collapsed to remove PCR duplicates, followed by removal of random nucleotides (four on both 5' and 3' end of the reads) using fastx_trimmer (FASTX Toolkit 0.0.14). Reads aligning to rRNA sequences and other sources of contamination using a custom index were removed by Bowtie2 (v2.3.2). The remaining sequences were aligned to the human transcriptome (hg19) using STAR aligner (2.5.3a) using GTF annotation file Gencode v19 by retaining only the reads that uniquely mapped to the hg19 genome (parameters --outFilterMultimapNmax 1 --outFilterMismatchNmax 5 --outFilterMatchNmin 15 --alignSJoverhangMin 5 --seedSearchStartLmax 20 --outSJfilterOverhangMin 30 8 8 8 --quantMode TranscriptomeSAM). Transcriptome BAM files were converted to the bed format using BedTools bamToBed (v2.26.0) and bed files were then input into riboWaltz (v1.1.0) (Lauria et al., 2018), which we used for downstream quality control and P site coverage analysis.

Quality control of ribosome profiling data was carried out using riboWaltz, which outputs metaheatmaps of read coverage around start and stop codons for all possible read lengths. Additionally, we used riboWaltz to calculate optimal P site offset (13 nt) and to obtain P site coverage per nucleotide. To include a P site in the downstream metagene analysis we requested a minimal P site nucleotide coverage of 5 in at least one of the samples.

To perform the metagene analysis of P site occupancy around the start and stop codons, we excluded the P site coverage corresponding to the first 2 and the last 2 codons of the CDS. The reason for this is the high read coverage at these positions due to ribosome initiation and termination. Including these would prevent meaningful normalization and mask the differences in the regions around start and stop codons. P site coverage per nucleotide was normalized for library size and summed for each codon. The codon coverage was scaled to the mean CDS coverage excluding the coverage at the extremities and we included all CDS that had total codon coverage of 5 or higher. For all interrogated

codon positions within the CDS, the scaled P site coverage was averaged and a rolling mean over 10 nt was used to smooth the signal in the final metagene plot.

In order to perform the metagene analysis of codon occupancy around targeting signals (Figs.5C,S5D), we summed the P site coverage per codon. The obtained codon coverage was normalized to the mean coverage within codons 20-40 in each CDS, as described previously (Chartron et al., 2016). For all interrogated codon positions within the CDS (1-500), the scaled P site coverage was next averaged over all well-quantified transcripts (mRNAs with TPM>=10) and a rolling mean over 5 nt was used to smooth the signal in the final metagene plot.

To calculate translation efficiencies per gene, we used RSEM (v1.2.2) which supplied us with read counts and TPM values per gene. The differences in translational efficiency, as well as in mRNA abundance and due to both effects (transcription and translation) were detected by DESeq2 (1.18.1) with an interaction term model as described previously (Chothani et al., 2019). Briefly, RPF read counts were normalized using the DESeq2 estimateSizeFactors function by taking into account all read counts and DESeq2 was run with default parameters. The log2-transformed fold changes for downstream comparisons were taken directly from the DESeq2 output.

For codon level analysis, we made use the codon frequency analysis available within riboWaltz. Differences in codon frequency were calculated by subtracting the codon occupancy values between conditions and additionally standard deviations between replicates were calculated.

Pulsed SILAC and data analysis

Experimental conditions

HEK293 WT and HDLBP knockout cell lines were cultured for at least 3 passages in arginine- and lysine-free DMEM (Life Technologies) containing 10% dialyzed FBS (Pan-Biotech), 1% glutamax (Life Technologies), 1% sodium pyruvate (Life Technologies) and “light” form amino acids 0,2 mM L-arginine (Sigma-Aldrich) and 0,8 mM L-lysine (Sigma-Aldrich). Next, cells were seeded in 6-well plates (450000 cells per well) and after 48 h the “light” form medium was removed. Subsequently, it was replaced by either medium containing the “medium” form amino acids (L-[13C6]-arginine (Sigma-Aldrich), L-[2H4]-lysine (Cambridge Isotope Laboratories)) or “heavy” form amino acids (L-[13C6,15N4]-arginine (Sigma-Aldrich), L-[13C6,15N2]-lysine (Cambridge Isotope Laboratories)). After 2 h or 4 h cells were fractionated. Cell fractionation was performed as described above with reduced volumes: 6-wells were washed with cold PBS and 544µl PBS containing 50 µg/ml cycloheximide was added for 10 min. In the meantime, cells were scraped and one 6-well HEK293 and one 6-well HEK293 HDLBP knockout cells were combined. Downstream we used 100 µl of permeabilization buffer, 544 µl of wash buffer and 100 µl of lysis buffer. After clarification of the

fractions, we recovered 90 µl sample, added 810 µl pure EtOH and samples were submitted to mass spectrometry.

Mass spectrometry

Samples were resuspended in 6 M urea, 2 M Thiourea, 10 mM HEPES pH 8 solution and the proteins were reduced with 10 mM DTT and alkylated with 55 mM iodoacetamide at room temperature. For lysis, we incubated the proteins with 1% (w/w) lysyl endopeptidase (Wako) at room temperature for 3 h, diluted with 50 mM ammonium bicarbonate solution for final urea concentration of 2 M and incubated with 1% (w/w) trypsin (Promega) under constant agitation at room temperature for 16 h. Next, peptides were acidified with 1% (v/v) trifluoroacetic acid, and desalted with C18 Stage Tips (Rappsilber et al., 2003). Prior to LC-MS/MS analysis, the peptides were eluted with 50% acetonitrile 0,1% formic acid, dried and resuspended in 3% acetonitrile, 0,1% formic acid (Buffer A). Peptide concentration was determined based on 280 nm UV light absorbance.

Reversed-phase liquid chromatography was carried out employing an EASY nLC II (Thermo Fisher Scientific) using self-made C18 microcolumns (75 µm ID, packed with ReproSil-Pur C18-AQ 1.9-µm resin, Dr. Maisch, Germany) connected on-line to the electrospray ion source (Proxeon, Denmark) of both an Orbitrao HF-X or an OrbitrapExploris 480 mass spectrometer with the FAIMS module installed and in application mode "Peptide" (Thermo Fisher Scientific). The peptides were eluted at a flow rate of 250 nL/min over 2 or 4 h with a stepwise increasing gradient of 4,74% to 81,3% acetonitrile in constant 0,1% formic acid. As settings for data dependent analysis on Orbitrap HF-X we used: positive polarity, full scan (resolution, 60000; m/z, 350-1800, AGC target, 3e6; injection time, 10 ms) followed by top20 MS/MS scans with higher-energy collisional dissociation (resolution, 15000; m/z, 200-2000; AGC target, 1e5. injection time, 22 ms; isolation width, 1.3 m/z; normalized collision energy, 26). The settings for data dependent analysis on Exploris were: positive polarity, 1 second cycle time, full scan (resolution, 60000; m/z, 350-1800, AGC target, 300%; injection time, 30 ms; FAIMS CV, -40, -55 or -70 mV) followed by MS/MS scans with higher-energy collisional dissociation (resolution, 7500; m/z, 200-2000; AGC target, 100%. injection time, 25 ms; isolation width, 1.3 m/z; normalized collision energy, 28). Ions with an unassigned charge state, singly charged ions, and ions with charge state higher than 5 were rejected and former target ions selected for MS/MS were dynamically excluded.

The samples were measured in technical duplicates. Raw files were analyzed using the MaxQuant software (v1.6.0.1) (Cox and Mann, 2008) with default parameters and files measured by Q-exactive HF-X or Exploris machines were grouped in same experimental group, while technical and biological replicates were kept separated. Search parameters included two missed cleavage sites, cysteine carbamidomethyl as fixed modification, and methionine oxidation, protein N-terminal acetylation

and asparagine or glutamine deamidation (only identification) as variable modifications. Triple multiplicity was used to search for light (Lys0, Arg0), medium-heavy (Lys4, Arg6) and heavy (Lys8, Arg10) peptides. The peptide mass tolerance was 20 and 4.5 ppm for first and main search, respectively. Database search was carried out with Andromeda embed in MaxQuant (Cox et al., 2011) against UniProt/Swiss-Prot human database (downloaded on January 2019) with common contaminant sequences provided by MaxQuant. FDR was set to 1% at peptide spectrum match (PSM) and protein levels. Minimum peptide count required for protein quantification was set to 2. Match between runs was turned on, search was performed with Requantify turned on, and again with option turned off.

Mass spectrometry proteomics data have been uploaded to ProteomeXchangeConsortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD018316.

Data analysis

For analysis, we excluded potential contaminants, reverse database hits and peptides only identified by modification. Next, unscrupulous ratios, defined as a SILAC pair quantified from both requantified intensities, were removed from further analysis. These were rare and only found in a small fraction of requantified SILAC ratios. The MaxQuant normalized SILAC ratios were used for analysis. For average calculations, we only accepted proteins with one or more values in at least one replicate from both forward and reverse SILAC label experiments. For the forward experiment, the heavy SILAC label corresponded to the HEK293 HDLBP KO cells, while the medium SILAC label corresponded to WT cells. For the reverse experiment we switched the labels. We noticed that the samples from the membrane fractions after 4 h of gradient fractionation contained higher iBAQ values for SP and TM containing proteins. Therefore, we only used membrane fraction results after 4 h for further analysis and excluded the 2 h data set. Protein names from the MaxQuant output were mapped to the RNA-sequencing fractionation data table to obtain the same classification as in other analyses. Distributions of H/M SILAC ratios between different classes of proteins were compared by cumulative density function. Significance was evaluated by the Wilcoxon rank-sum test.

Polysome profiling

Polysome profiling was performed as described previously (Jagannathan et al., 2011). Stable cell lines expressing FLAG/HA-HDLBP (4 dishes (15 cm) per replicate) were grown to 90% confluency. 4 ml ice cold PBS containing 100 µg/ml cycloheximide per plate was added and cells were scraped using a plastic scraper. After pelleting the cells (1200 rpm, 4 min, 4°C), cells were resuspended with 250 µl polysome lysis buffer (20 mM Tris-HCl [pH 7.4], 150 mM KCl, 5 mM $MgCl_2$, 0,5 mM DTT, 0,5% IGEPAL

CA-630, EDTA-free protease inhibitor cocktail [Roche], 100 µg/ml cycloheximide). The lysate was passed for 4 min through a 26 G needle and centrifuged (17000 g, 10 min, 4°C). The supernatant was layered onto a 10 ml linear sucrose gradient (15-45%) using a Sorvall WX 90 ultracentrifuge (Thermo Fisher Scientific) at 36000 rpm at 4°C for 2 h 30 min. The sucrose gradients were prepared by first adding 5 ml of 15% sucrose solution (15% sucrose in polysome lysis buffer (without DTT, EDTA-free protease inhibitor, cycloheximide) in a centrifuge tube (Seton, 7030) and careful under laying of 45% sucrose solution. The gradient station (Biocomp) was used to prepare the gradients and to collect 20 fractions per gradient. Proteins from each fraction were extracted by standard TCA precipitation and analyzed by Western Blot.

Related to results chapter [9]

ncRNA-sequencing and data analysis

Experimental conditions

For normalization of tRNA abundance in the HDLBP PAR-CLIP dataset we used the RNA-sequencing library preparation protocol based on small RNA cloning approaches (Hafner et al., 2008; König et al., 2010; Milek et al., 2017; Van Nostrand et al., 2016). Total RNA was extracted from HEK293 cells with Qiagen RNAeasy kit according to the manufacturer's protocol and 2 µg were fragmented in a total volume of 100 µl fragmentation solution (10 mM ZnOAc, 100 mM Tris buffer [pH 8.0]) in a thin-walled PCR tube at 95°C for 55 s. Next, the samples were immediately cooled on ice. The RNA was purified with RNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's protocol. RNA was treated with 10 U of calf intestine alkaline phosphatase in 1x CutSmart buffer (both from New England Biolabs) for 1 h at 37 °C. The samples were immediately cooled on ice and RNA was purified with RNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's protocol. RNA was then treated with 20 U of T4 polynucleotide kinase in 1x T4 PNK reaction buffer (both from New England Biolabs) supplemented with 0,1% (v/v) Triton-X100 for 30 min at 37°C. The samples were immediately cooled on ice and RNA was purified with RNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's protocol. Next, the 3' adaptor was ligated, which was performed for 16 h at 16 °C in a total volume of 20 µl in the presence of 2 U of T4 RNA ligase 2 truncated K227Q, 1x RNA ligase buffer, 25% PEG-8000 (all from New England Biolabs) and 25 pmol of pre-adenylated 4N-RA3 adapter (see "Oligonucleotides"). Afterwards, RNA was separated on 10% Novex TBE Urea gels. Fragments in the range of 65-100 nt were excised and eluted in 300 mM NaCl overnight at 4 °C with 900 rpm agitation. RNA was then precipitated by the addition of 0,5 µl GlycoBlue co-precipitant, 3 volumes of absolute ethanol and incubation for at least 2 h at -

80 °C. The precipitated RNA was reverse transcribed for 30 min at 37 °C in a total volume of 20 µl containing 1x buffer (25 mM Tris [pH 8.6], 30 mM NaCl and 5 mM dithiotreitol), 4.5 pmol RTP primer (see "Oligonucleotides"), 4 mM each dNTP, 20 mM $MgCl_2$, 7.1% (v/v) dimethylsulfoxide and 10 U of AMV reverse transcriptase (Promega). Subsequently, RNA was removed by RNase H (5 U) treatment for 30 min at 37 °C (New England Biolabs) and ssDNA was separated on 10% Novex TBE Urea gels. Fragments ranging from 65-100 nt were excised and eluted in 300 mM NaCl overnight at 4 °C with 900 rpm agitation. Single stranded DNA (ssDNA) was then precipitated by the addition of 0,5 µl GlycoBlue co-precipitant, 3 volumes of absolute ethanol and incubation for at least 2 h at -80 °C. Next, the 5'adaptor was ligated, which was performed for 16 h at 22 °C in a total volume of 20 µl. The reaction contained precipitated ssDNA, 60 U T4 RNA ligase 1, 1x RNA ligase buffer, 22.5% PEG-8000, 1 mM ATP (all from New England Biolabs) and 50 pmol of 4N-SRC-cDNA adapter (see "Oligonucleotides"). The ssDNA was purified with DNA Clean and Concentrator kit (Zymo Research) according to the manufacturer's ssDNA protocol. PCR amplification for the final library was performed according to the Truseq small RNA protocol (Illumina). Typically, 14 cycles of amplification were necessary to obtain a sufficient amount of the final cDNA library in the size range of 150-220 bp (average size 180 bp). PCR reactions were separated on a 2.5% agarose gel. The final DNA library of the expected size range was excised and sequenced on a HiSeq 4000 Illumina instrument (1x51+7 cycle).

Data analysis

Reads were demultiplexed and adapter sequences were removed by Flexbar (2.5). Reads were collapsed to remove PCR duplicates, followed by removal of random nucleotides (four on both 5' and 3' end of the reads) by fastx_trimmer (FASTX Toolkit 0.0.14). To obtain accurate mapping results to transcripts originating from repetitive genomic loci, we made use of the same strategy as for PAR-CLIP analysis (see "PAR-CLIP and data analysis").

Mapping to reference tRNAs and quantification

Initially, we downloaded mature tRNA fasta files from GtRNADb (http://gtrnadb.ucsc.edu/, hg19 build) and kept only one fasta sequence per tRNA molecule to avoid mapping non-uniquely due to tRNA gene duplications and pseudogenes. The tRNAs missing from this initial set were obtained from HEK293 Hydro-seq results (Gogakos et al., 2017). As a result our custom made FASTA tRNA reference contained 58 unique tRNAs. Reads were mapped to this reference using Bowtie 2 (v.2.3.2) and custom scripts using BedTools (v2.26.0) intersect and getfasta commands were used to obtain T-to-C conversion counts per position and per tRNA. By inspection of the aligned reads in IGV and using the Modomics database (https://iimcb.genesilico.pl/modomics) (Boccaletto et al., 2018), we realized

that T-to-C conversions were detected at dihydrouridine bases, which were present also in libraries prepared from total HEK293 RNA with no 4sU and UV treatment. Therefore, we removed these positions from T-to-C conversion counting by using a custom made bed file of all dihydrouridine bases. To quantify the tRNA abundance in HEK293 cells we used our ncRNA-seq dataset from randomly fragmented total HEK293 RNA (see "ncRNA-sequencing and data analysis"). To calculate the enrichment of tRNAs in HDLBP PAR-CLIP, we first normalized T-to-C conversion counts and read counts per tRNA for library size. Next, for each tRNA, T-to-C conversion counts were divided by the respective read count per tRNA from the ncRNA-seq samples to obtain the enrichment score. Transition specificity per T-to-C conversion within one tRNA was obtained by dividing the T-to-C conversion count with total read coverage at the same nucleotide position. Codon usage information was downloaded from https://www.kazusa.or.jp/codon/. The dihydrouridine positions along with the entire tRNA processing pipeline and tRNA sequences are available from GitHub https://github.com/mmilek/hdlbp/trna/.

Related to results chapter [10]

***Gaussia* luciferase and SEAP assays**

HEK293 WT and HDLBP knockout cell lines were seeded in 6-well plates (500000 cells per well). After 24 h cells were transfected using Fugene6 reagent (Promega) with plasmids expressing *Gaussia* luciferase (Gluc) or secreted alkaline phosphatase (SEAP) and Firefly luciferase (Fluc) according to manufacturer´s protocol (see "Plasmids" for details). 48 h after transfection cells were split (70000 cells per well) into white bottom 96-well plates (Thermo Fisher Scientific, 136101). 24 h later a proportion of medium was transferred to a new white bottom 96-well plate. The Gluc signal was measured using the *Gaussia* Luciferase Flash Assay Kit (Thermo Fisher Scientific, PI16159) and the SEAP signal was quantified by the NovaBrightPhospha-Light EXP Assay Kit (Thermo Fisher Scientific, N10577) according to manufacturer´s protocol. The Fluc activity was determined in the remaining medium and cells with the Firefly Luc One-Step Glow Assay Kit (Thermo Fisher Scientific, PI16197). Gluc and SEAP signal was normalized to the Fluc signal. Each experiment was carried out using 5 technical replicates and the Gluc assay was performed 5 times and the SEAP assay 4 times.

For SEAP rescue experiments we transfected HEK293 WT and HDLBP KO gRNA2 cells with SEAP, Fluc and additionally with pFRT/TO/FLAG/HA-HDLBP for 24 h. Subsequent steps of the experiment were carried out as described above.

Flow cytometry

To quantify changes in CD71 surface expression, $3x10^5$ HEK293 WT and HDLBP KO cells were stained with an anti-CD71 antibody or an isotype control diluted in 1% BSA/PBS for 15 min at 4°C. After washing the cells with PBS, cells were analyzed by flow cytometry using an Aria III (BD) machine. To exclude dead cells, cells were stained with propidium iodid (Miltenyi Biotec). The FlowJo analysis software was used to determine mean fluorescence intensities of CD71. The experiment was carried out 4 times.

Related to results chapter [11]

Cell Proliferation and Migration Assays

For the determination of cell proliferation, 10^3 A549 WT or HDLBP KO cells were seeded in 96-well plates. Cell confluency was monitored for 5 days using an IncuCyte S3 system (Sartorius) with 4x magnification to allow a whole well scan. Confluence masks were obtained by the IncuCyte analysis software. Cell viability and DNA content were quantified using CellTiter Glo (Promega) supplemented with 1/2000 SYBR Green I (Thermo Fisher) according to manufacturer's protocols. Luminescence and fluorescence intensities were determined in a GloMax microplate reader (Promega). For the scratch wound migration analysis, $2.5x10^4$ A549 WT and HDLBP KO cells were seeded in a 96-well ImageLock plate (Sartorius) for 24 h. The wound areas were created on confluent cell monolayers by a 96-well WoundMaker (Sartorius). Wound healing was monitored in 1 h intervals using an IncuCyte S3 system at 10x magnification. Confluence and wound masks were obtained and quantified by the IncuCyte analysis software.

Xenograft Assay

Athymic nude mice (immunodeficient, $FOXN1^{nu/nu}$) were purchased from Charles River. Animals were treated according to the local guidelines of the Martin-Luther-University Halle-Wittenberg. Subcutaneous xenograft assays were carried out as previously described (Muller et al., 2018). Briefly, A549 WT and HDLBP KO cells were transduced at 10 MOI (multiplicity of injection) with iRFP-encoding lentiviruses. $1x10^6$ cells were harvested in media supplemented with 50% (v/v) matrigel (Sigma Aldrich) for the subcutaneous injection into nude mice. WT and HDLBP KO cells were injected into the left and the right flanks of six-week old nude mice. Mice were held with access to chlorophyll-free food (Altromin C1086) *ad libitum* in order to reduce background noise in weekly iRFP imaging using a Pearl Trilogy Imaging System (LI-COR). To determine the fluorescence intensity, we used the Image Studio software (LI-COR). Tumor volumes were measured and calculated using the formula

0.52 x L_1 x L_2 x L_3. Once the first tumor reached the termination criteria of a 1.5 cm diameter, the mice were sacrificed. Palpable tumors were excised and the weight was measured. Subsequently, the RNA was isolated and sequencing libraries were prepared using the Next Ultra Directional RNA Library Prep Kit (NEB). Next generation sequencing performed on a HiSeq 4000 Illumina instrument (1x51+7 cycles). The permission for *in vivo* xenograft assays was granted by an ethical review committee.

General information

Data and code availability

PAR-CLIP, RNA-seq and ribosome profiling data from this study have been uploaded to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession GSE148262. BioID and pSILAC raw data have been submitted to the ProteomeXchange (http://www.proteomexchange.org) under the identifiers PXD018313 and PXD018316 respectively. All processed data and analysis scripts are available from GitHub at https://github.com/mmilek/hdlbp.git.

Antibodies

Antibody	*Vendor*	*Dilution*
APC-isotype control	Miltenyi Biotec RRID:AB_2733447	1 : 50
BCAP-31	Proteintech, 11200-1-AP	1 : 2000
CD71-APC	Miltenyi Biotec RRID:AB_2660542	1 : 50
Goat Anti-Mouse Immunoglobulins/HRP	Agilent-Dako, P0447	1 : 4000
Goat Anti-Rabbit Alexa Fluor 488	Invitrogen, A-11008	1 : 2000
Goat Anti-Rabbit Immunoglobulins/HRP	Agilent-Dako, P0448	1 : 4000
HA	Covance, MMS-101P-1000	1 : 5000
HDLBP	Abcam, ab109324	1 : 10 000
RPL7	Abcam, ab72550	1 : 5000
RPS6	Cell Signaling, #2217	1 : 1000
Tubulin	Sigma-Aldrich, T8328	1 : 4000

Oligonucleotides

Oligonucleotide	*Sequence*
27 nt marker	5′-AUGUACACGGAGUCGAGCUCAACCCGC-P
30 nt marker	5′-AUGUACACGGAGUCGAGCUCAACCCGCAAC-P
3' adaptor 4N-RA3	5'-rApp/NNNNTGGAATTCTCGGGTGCCAAGG/3InvdT/
5' adaptor OR5-4N	5'-GUUCAGAGUUCUACAGUCCGACGAUCNNNN
5'adaptor 4N-SRC-cDNA	/5Phos/NNNNGATCGTCGGACTGTAGAACTCTGAAC-SpC3
RT primer RTP	5' GCCTTGGCACCCGAGAATTCCA
RP1 (forward)	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA
RPI1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RPI8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
*Gaussia*_fwd	AACTGCTAGCATGGGAGTCAAAGTTCTGTTTG
*Gaussia*_rev	AATTGCGGCCGCTTAGTCACCACCG
SEAP_fwd	AACTGCTAGCATGCTGCTGCTGCTGCTGCTGCTGG
SEAP_rev	AATTGCGGCCGCCTAGCCGGCCGCCCCGACTCTAGA
ATP1A1_fwd	GGCAGTGTTTCAGGCTAACCAG
ATP1A1_rev	TCTCCTTCACGGAACCACAGCA
ATP1A3_fwd	GTGGACATACGAGCAGAGGA
ATP1A3_rev	ACGATGCTCACAAAGAAGGC
CD46_fwd	TGGCTACCTGTCTCAGATGACG
CD46_rev	GCATCTGATAACCAAACTCGTAAG
CD71_fwd	ATCGGTTGGTGCCACTGAATGG
CD71_rev	ACAACAGTGGGCTGGCAGAAAC
HNRNPUL1_fwd	GCAGAGAACGATGTGATTGGCTG
HNRNPUL1_rev	CAAGGCTTCCTTCTGGATTCGG
IGF2R_fwd	CGACTGCCAGTACCTCTTCT
IGF2R_rev	CGGGATTCTCGCTGTCAAAG
LAMB1_fwd	ACATTGACACGACAGACCC
LAMB1_rev	GCATCACCATAGTATCCAAACC
LAMC1_fwd	GGACATTCGCAATCTGGAGG
LAMC1_rev	CAATGGACGGGGTGTTGAAG
MT-CO1_fwd	GACGTAGACACACGAGCATATTTCA
MT-CO1_rev	AGGACATAGTGGAAGTGAGCTACAAC
PFN1_fwd	CATCGTGGGCTACAAGGACTCG
PFN1_rev	CCAAGTGTCAGCCCATTCACGT
RN7SL_fwd	TCAATATGGTGACCTCCCGG
RN7SL_rev	TTTTGACCTGCTCCGTTTCC
TMBIM6_fwd	CTGTATGTGGGACTGGTGGT
TMBIM6_rev	TGATCTCCATGTTCGGCCTT
YWHAZ_fwd	ACCGTTACTTGGCTGAGGTTGC
YWHAZ_rev	CCCAGTCTGATAGGATGTGTTGG

Plasmids

Vector pDONR221 carrying the HDLBP CDS (obtained from the hORFeome V5.1 collection) was recombined into pFRT/TO/FLAG/HA-DEST (Addgene ID: 26360) and pDEST5-BirA-FLAG-N-term-pcDNA5-FRT-TO (Couzens et al., 2013) using the Gateway LR Clonase II (Thermo Fischer Scientific) according to the manufacturer´s protocol. For the SEAP secretion assays we made use of pEZX-GA01

(GeneCopoeia) and additionally pFRTpsiCHECK containing Renilla luciferase and Firefly luciferase. For the SEAP rescue secretion assays we replaced Renilla luciferase in pFRTpsiCHECK by SEAP luciferase. Therefore, we amplified SEAP from pEZX-GA01 with the primers SEAP_fwd and SEAP_rev (see "Oligonucleotides") and ligation into NheI and NotI restriction sites. For the *Gaussia* luciferase secretion assay, we replaced Renilla luciferase in pFRTpsiCHECK by *Gaussia* luciferase. Therefore, we amplified *Gaussia* from pEZX-GA01 with the primers *Gaussia*_fwd and *Gaussia*_rev (see "Oligonucleotides") and ligation into NheI and NotI restriction sites. All vectors have been submitted to Addgene.

REFERENCES

Acosta-Alvear, D., Karagöz, G.E., Fröhlich, F., Li, H., Walther, T.C., and Walter, P. (2018). The unfolded protein response and endoplasmic reticulum protein targeting machineries converge on the stress sensor IRE1. eLife *7*.

Anger, A.M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature *497*, 80–85.

Aviram, N., and Schuldiner, M. (2017). Targeting and translocation of proteins to the endoplasmic reticulum at a glance. J. Cell Sci. *130*, 4079–4085.

Aviram, N., Ast, T., Costa, E.A., Arakel, E.C., Chuartzman, S.G., Jan, C.H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. Nature *540*, 134–138.

Barchiesi, A., and Vascotto, C. (2019). Transcription, Processing, and Decay of Mitochondrial RNA in Health and Disease. Int. J. Mol. Sci. *20*.

Baum, S., Bittins, M., Frey, S., and Seedorf, M. (2004). Asc1p, a WD40-domain containing adaptor protein, is required for the interaction of the RNA-binding protein Scp160p with polysomes. Biochem. J. *380*, 823–830.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

Bernstein, H.D., Zopf, D., Freymann, D.M., and Walter, P. (1993). Functional substitution of the signal recognition particle 54-kDa subunit by its Escherichia coli homolog. Proc. Natl. Acad. Sci. U. S. A. *90*, 5229–5233.

Biswas, J., Nunez, L., Das, S., Yoon, Y.J., Eliscovich, C., and Singer, R.H. (2019). Zipcode Binding Protein 1 (ZBP1; IGF2BP1): A Model for Sequence-Specific RNA Regulation. Cold Spring Harb. Symp. Quant. Biol. *84*, 1–10.

Blobel, G., and Sabatini, D.D. (1971). Ribosome-membrane interaction in eukaryotic cells. In Biomembranes, (Springer), pp. 193–195.

Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092–1096.

Boccaletto, P., Machnicka, M.A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T.K., de Crécy-Lagard, V., Ross, R., Limbach, P.A., Kotter, A., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res. *46*, D303–D307.

Brykailo, M.A., Corbett, A.H., and Fridovich-Keil, J.L. (2007). Functional overlap between conserved and diverged KH domains in Saccharomyces cerevisiae SCP160. Nucleic Acids Res. *35*, 1108–1118.

Buskirk, A.R., and Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and eukaryotes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *372*.

Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. *44*, D1251–D1257.

Cha, K., Li, Y., and Yi, G.-S. (2016). Discovering gene expression signatures responding to tyrosine kinase inhibitor treatment in chronic myeloid leukemia. BMC Med. Genomics *9 Suppl 1*, 29.

Chartron, J.W., Hunt, K.C.L., and Frydman, J. (2016). Cotranslational signal-independent SRP preloading during membrane targeting. Nature *536*, 224–228.

Cheng, M.H., and Jansen, R.-P. (2017). A jack of all trades: the RNA-binding protein vigilin: Vigilin: a multifunctional RNA-binding protein. Wiley Interdiscip. Rev. RNA e1448.

Cheng, M.H.K., Hoffmann, P.C., Franz-Wachtel, M., Sparn, C., Seng, C., Maček, B., and Jansen, R.-P. (2018). The RNA-Binding Protein Scp160p Facilitates Aggregation of Many Endogenous Q/N-Rich Proteins. Cell Rep. *24*, 20–26.

Cho, J., Chang, H., Kwon, S.C., Kim, B., Kim, Y., Choe, J., Ha, M., Kim, Y.K., and Kim, V.N. (2012). LIN28A is a suppressor of ER-associated translation in embryonic stem cells. Cell *151*, 765–777.

Chothani, S., Adami, E., Ouyang, J.F., Viswanathan, S., Hubner, N., Cook, S.A., Schafer, S., and Rackham, O.J.L. (2019). deltaTE: Detection of Translationally Regulated Genes by Integrative Analysis of Ribo-seq and RNA-seq Data. Curr. Protoc. Mol. Biol. *129*, e108.

Cohen-Zontag, O., Baez, C., Lim, L.Q.J., Olender, T., Schirman, D., Dahary, D., Pilpel, Y., and Gerst, J.E. (2019). A secretion-enhancing cis regulatory targeting element (SECReTE) involved in mRNA localization and protein synthesis. PLoS Genet. *15*, e1008248.

Costa, E.A., Subramanian, K., Nunnari, J., and Weissman, J.S. (2018). Defining the physiological role of SRP in protein-targeting efficiency and specificity. Science *359*, 689–692.

Couzens, A.L., Knight, J.D.R., Kean, M.J., Teo, G., Weiss, A., Dunham, W.H., Lin, Z.-Y., Bagshaw, R.D., Sicheri, F., Pawson, T., et al. (2013). Protein Interaction Network of the Mammalian Hippo Pathway Reveals Mechanisms of Kinase-Phosphatase Interactions. Sci. Signal. *6*, rs15–rs15.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Cui, X.A., and Palazzo, A.F. (2014). Localization of mRNAs to the endoplasmic reticulum. Wiley Interdiscip. Rev. RNA *5*, 481–492.

Cui, X.A., Zhang, Y., Hong, S.J., and Palazzo, A.F. (2013). Identification of a region within the placental alkaline phosphatase mRNA that mediates p180-dependent targeting to the endoplasmic reticulum. J. Biol. Chem. *288*, 29633–29641.

Dagil, R., Ball, N.J., Ogrodowicz, R.W., Hobor, F., Purkiss, A.G., Kelly, G., Martin, S.R., Taylor, I.A., and Ramos, A. (2019). IMP1 KH1 and KH2 domains create a structural platform with unique RNA recognition and re-modelling properties. Nucleic Acids Res. *47*, 4334–4348.

Díaz-López, I., Toribio, R., Berlanga, J.J., and Ventoso, I. (2019). An mRNA-binding channel in the ES6S region of the translation 48S-PIC promotes RNA unwinding and scanning. eLife *8*.

Eliseev, B., Yeramala, L., Leitner, A., Karuppasamy, M., Raimondeau, E., Huard, K., Alkalaeva, E., Aebersold, R., and Schaffitzel, C. (2018). Structure of a human cap-dependent 48S translation pre-initiation complex. Nucleic Acids Res. *46*, 2678–2689.

Felder, B., Radlwimmer, B., Benner, A., Mincheva, A., Tödt, G., Beyer, K.S., Schuster, C., Bölte, S., Schmötzer, G., Klauck, S.M., et al. (2009). FARP2, HDLBP and PASK are downregulated in a patient with autism and 2q37.3 deletion syndrome. Am. J. Med. Genet. A. *149A*, 952–959.

Frey, S., Pool, M., and Seedorf, M. (2001). Scp160p, an RNA-binding, polysome-associated protein, localizes to the endoplasmic reticulum of Saccharomyces cerevisiae in a microtubule-dependent manner. J. Biol. Chem. *276*, 15905–15912.

Garzia, A., Meyer, C., Morozov, P., Sajek, M., and Tuschl, T. (2017). Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. Methods San Diego Calif *118-119*, 24–40.

Gelin-Licht, R., Paliwal, S., Conlon, P., Levchenko, A., and Gerst, J.E. (2012). Scp160-dependent mRNA trafficking mediates pheromone gradient sensing and chemotropism in yeast. Cell Rep. *1*, 483–494.

Genuth, N.R., and Barna, M. (2018). The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. Mol. Cell *71*, 364–374.

Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M., and Tuschl, T. (2017). Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. Cell Rep. *20*, 1463–1475.

Graham, D.L., and Oram, J.F. (1987). Identification and characterization of a high density lipoprotein-binding protein in cell membranes by ligand blotting. J. Biol. Chem. *262*, 7439–7442.

Gregersen, L.H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. Mol. Cell *54*, 573–585.

Grishin, N.V. (2001). KH domain: one motif, two folds. Nucleic Acids Res. *29*, 638–643.

Guna, A., and Hegde, R.S. (2018). Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. Curr. Biol. CB *28*, R498–R511.

Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., and Tuschl, T. (2008). Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. Methods San Diego Calif *44*, 3–12.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

Halic, M., Becker, T., Pool, M.R., Spahn, C.M.T., Grassucci, R.A., Frank, J., and Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. Nature *427*, 808–814.

Hinnebusch, A.G. (2014). The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. *83*, 779–812.

Hirschmann, W.D., Westendorf, H., Mayer, A., Cannarozzi, G., Cramer, P., and Jansen, R.-P. (2014). Scp160p is required for translational efficiency of codon-optimized mRNAs in yeast. Nucleic Acids Res. *42*, 4043–4055.

Hsu, J.C.-C., and Nicchitta, C.V. (2018). Proteome expression: the subcellular organisation of protein synthesis. eLS 1–8.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science *324*, 218–223.

Jagannathan, S., Nwosu, C., and Nicchitta, C.V. (2011). Analyzing mRNA Localization to the Endoplasmic Reticulum via Cell Fractionation. In RNA Detection and Visualization, J.E. Gerst, ed. (Totowa, NJ: Humana Press), pp. 301–321.

Jagannathan, S., Reid, D.W., Cox, A.H., and Nicchitta, C.V. (2014). De novo translation initiation on membrane-bound ribosomes as a mechanism for localization of cytosolic protein mRNAs to the endoplasmic reticulum. RNA *20*, 1489–1498.

Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science *346*, 1257521–1257521.

Kanamori, H., Dodson, R.E., and Shapiro, D.J. (1998). In vitro genetic analysis of the RNA binding site of vigilin, a multi-KH-domain protein. Mol. Cell. Biol. *18*, 3991–4003.

Kim, Y.-I., Lee, J., Choi, Y.-J., Seo, J., Park, J., Lee, S.-Y., and Cho, J.-Y. (2015). Proteogenomic Study beyond Chromosome 9: New Insight into Expressed Variant Proteome and Transcriptome in Human Lung Adenocarcinoma Tissues. J. Proteome Res. *14*, 5007–5016.

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat. Struct. Mol. Biol. *17*, 909–915.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Kügler, S., Grünweller, A., Probst, C., Klinger, M., Müller, P.K., and Kruse, C. (1996). Vigilin contains a functional nuclear localisation sequence and is present in both the cytoplasm and the nucleus. FEBS Lett. *382*, 330–334.

Lakkaraju, A.K.K., Luyet, P.-P., Parone, P., Falguières, T., and Strub, K. (2007). Inefficient targeting to the endoplasmic reticulum by the signal recognition particle elicits selective defects in post-ER membrane trafficking. Exp. Cell Res. *313*, 834–847.

Lang, B.D., and Fridovich-Keil, J.L. (2000). Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. Nucleic Acids Res. *28*, 1576–1584.

Lang, B.D., Li Am, null, Black-Brewster, H.D., and Fridovich-Keil, J.L. (2001). The brefeldin A resistance protein Bfr1p is a component of polyribosome-associated mRNP complexes in yeast. Nucleic Acids Res. *29*, 2567–2574.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. Mol. Cell *43*, 340–352.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, A., Vargas, C.A., Brykailo, M.A., Openo, K.K., Corbett, A.H., and Fridovich-Keil, J.L. (2004). Both KH and non-KH domain sequences are required for polyribosome association of Scp160p in yeast. Nucleic Acids Res. *32*, 4768–4775.

Li, A.-M., Watson, A., and Fridovich-Keil, J.L. (2003). Scp160p associates with specific mRNAs in yeast. Nucleic Acids Res. *31*, 1830–1837.

Li, J., Ni, M., Lee, B., Barron, E., Hinton, D.R., and Lee, A.S. (2008). The unfolded protein response regulator GRP78/BiP is required for endoplasmic reticulum integrity and stress-induced autophagy in mammalian cells. Cell Death Differ. *15*, 1460–1471.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

Loya, A., Pnueli, L., Yosefzon, Y., Wexler, Y., Ziv-Ukelson, M., and Arava, Y. (2008). The 3'-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. RNA N. Y. N *14*, 1352–1365.

Lu, S.H.-J., Jeon, A.H.W., Schmitt-Ulms, G., Qamar, S., Dodd, R., McDonald, B., Li, Y., Meadows, W., Cox, K., Bohm, C., et al. (2012). Vigilin interacts with signal peptide peptidase. Proteome Sci. *10*, 33.

Maquat, L.E., Tarn, W.-Y., and Isken, O. (2010). The pioneer round of translation: features and functions. Cell *142*, 368–374.

McCaffrey, K., and Braakman, I. (2016). Protein quality control at the endoplasmic reticulum. Essays Biochem. *60*, 227–235.

Milek, M., Imami, K., Mukherjee, N., Bortoli, F.D., Zinnall, U., Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., et al. (2017). DDX54 regulates transcriptome dynamics during DNA damage response. Genome Res. *27*, 1344–1359.

Mobin, M.B., Gerstberger, S., Teupser, D., Campana, B., Charisse, K., Heim, M.H., Manoharan, M., Tuschl, T., and Stoffel, M. (2016). The RNA-binding protein vigilin regulates VLDL secretion through modulation of Apob mRNA translation. Nat. Commun. *7*, 12848.

Molyneux, S.D., Waterhouse, P.D., Shelton, D., Shao, Y.W., Watling, C.M., Tang, Q.-L., Harris, I.S., Dickson, B.C., Tharmapalan, P., Sandve, G.K., et al. (2014). Human somatic cell mutagenesis creates genetically tractable sarcomas. Nat. Genet. *46*, 964–972.

Mueckler, M.M., and Pitot, H.C. (1982). Structure and function of rat liver polysome populations. II. Characterization of polyadenylate-containing mRNA associated with subpopulations of membrane-bound particles. J. Cell Biol. *94*, 297–307.

Mullari, M., Lyon, D., Jensen, L.J., and Nielsen, M.L. (2017). Specifying RNA-Binding Regions in Proteins by Peptide Cross-Linking and Affinity Purification. J. Proteome Res. *16*, 2762–2772.

Mutka, S.C., and Walter, P. (2001). Multifaceted physiological response allows yeast to adapt to the loss of the signal recognition particle-dependent protein-targeting pathway. Mol. Biol. Cell *12*, 577–588.

Neu-Yilik, G., Zorbas, H., Gloe, T.R., Raabe, H.M., Hopp-Christensen, T.A., and Müller, P.K. (1993). Vigilin is a cytoplasmic protein. A study on its expression in primary cells and in established cell lines of different species. Eur. J. Biochem. FEBS *213*, 727–736.

Nicastro, G., Taylor, I.A., and Ramos, A. (2015). KH-RNA interactions: back in the groove. Curr. Opin. Struct. Biol. *30*, 63–70.

Ooi, Y.S., Majzoub, K., Flynn, R.A., Mata, M.A., Diep, J., Li, J.K., van Buuren, N., Rumachik, N., Johnson, A.G., Puschnik, A.S., et al. (2019). An RNA-centric dissection of host complexes controlling flavivirus infection. Nat. Microbiol.

Palade, G.E. (1955). A small particulate component of the cytoplasm. J. Biophys. Biochem. Cytol. *1*, 59–68.

Palade, G.E. (1966), ER electron microscopy doi:10.7295/W9CIL37237

Pechmann, S., Chartron, J.W., and Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. Nat. Struct. Mol. Biol. *21*, 1100–1105.

Pfeffer, S., Brandt, F., Hrabe, T., Lang, S., Eibauer, M., Zimmermann, R., and Förster, F. (2012). Structure and 3D arrangement of endoplasmic reticulum membrane-associated

ribosomes. Struct. Lond. Engl. 1993 *20*, 1508–1518.

Polyansky, A.A., Hlevnjak, M., and Zagrovic, B. (2013). Analogue encoding of physicochemical properties of proteins in their cognate messenger RNAs. Nat. Commun. *4*, 2784.

Prilusky, J., and Bibi, E. (2009). Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. Proc. Natl. Acad. Sci. U. S. A. *106*, 6662–6666.

Pyhtila, B., Zheng, T., Lager, P.J., Keene, J.D., Reedy, M.C., and Nicchitta, C.V. (2008). Signal sequence- and translation-independent mRNA localization to the endoplasmic reticulum. RNA *14*, 445–453.

Rabouille, C. (2017). Pathways of Unconventional Protein Secretion. Trends Cell Biol. *27*, 230–240.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663–670.

Reid, D.W., and Nicchitta, C.V. (2012). Primary Role for Endoplasmic Reticulum-bound Ribosomes in Cellular Translation Identified by Ribosome Profiling. J. Biol. Chem. *287*, 5518–5527.

Reid, D.W., and Nicchitta, C.V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. Nat. Rev. Neurosci.

Rodnina, M.V. (2016). The ribosome in action: Tuning of translational efficiency and protein folding. Protein Sci. Publ. Protein Soc. *25*, 1390–1406.

Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568–584.e23.

Schibich, D., Gloge, F., Pöhner, I., Björkholm, P., Wade, R.C., von Heijne, G., Bukau, B., and Kramer, G. (2016). Global profiling of SRP interaction with nascent polypeptides. Nature *536*, 219–223.

Schmidt, C., Henkel, B., Pöschl, E., Zorbas, H., Purschke, W.G., Gloe, T.R., and Müller, P.K. (1992). Complete cDNA sequence of chicken vigilin, a novel protein with amplified and evolutionary conserved domains. Eur. J. Biochem. FEBS *206*, 625–634.

Schueler, M., Munschauer, M., Gregersen, L.H., Finzel, A., Loewer, A., Chen, W., Landthaler, M., and Dieterich, C. (2014). Differential protein occupancy profiling of the mRNA transcriptome. Genome Biol. *15*, R15.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell *134*, 634–645.

Schuller, A.P., and Green, R. (2018). Roadblocks and resolutions in eukaryotic translation. Nat. Rev. Mol. Cell Biol. *19*, 526–541.

Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. (2009). Global analysis of cellular protein translation by pulsed SILAC. Proteomics *9*, 205–209.

Shibata, Y., Voeltz, G.K., and Rapoport, T.A. (2006). Rough sheets and smooth tubules. Cell *126*, 435–439.

Silva, L.A., and Dermody, T.S. (2017). Chikungunya virus: epidemiology, replication, disease mechanisms, and prospective intervention strategies. J. Clin. Invest. *127*, 737–749.

Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., Chang, H.Y., and Barna, M. (2017). The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. Cell *169*, 1051–1065.e18.

Stefanovic, S., and Hegde, R.S. (2007). Identification of a targeting factor for posttranslational membrane protein insertion into the ER. Cell *128*, 1147–1159.

Trendel, J., Schwarzl, T., Horos, R., Prakash, A., Bateman, A., Hentze, M.W., and Krijgsveld, J. (2019). The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. Cell *176*, 391–403.e19.

Valverde, R., Edwards, L., and Regan, L. (2008). Structure and function of KH domains. FEBS J. *275*, 2712–2726.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

Viotti, C. (2016). ER to Golgi-Dependent Protein Secretion: The Conventional Pathway. Methods Mol. Biol. Clifton NJ *1459*, 3–29.

Voigt, F., Zhang, H., Cui, X.A., Triebold, D., Liu, A.X., Eglinger, J., Lee, E.S., Chao, J.A., and Palazzo, A.F. (2017). Single-Molecule Quantification of Translation-Dependent Association of mRNAs with the Endoplasmic Reticulum. Cell Rep. *21*, 3740–3753.

Vollbrandt, T., Willkomm, D., Stossberg, H., and Kruse, C. (2004). Vigilin is co-localized with 80S ribosomes and binds to the ribosomal complex through its C-terminal domain. Int. J. Biochem. Cell Biol. *36*, 1306–1318.

Voorhees, R.M., and Hegde, R.S. (2015). Structures of the scanning and engaged states of the mammalian SRP-ribosome complex. eLife *4*.

Walter, P., and Blobel, G. (1980). Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. Proc. Natl. Acad. Sci. U. S. A. *77*, 7112–7116.

Walter, P., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. J. Cell Biol. *91*, 557–561.

Wang, M., and Kaufman, R.J. (2016). Protein misfolding in the endoplasmic reticulum as a conduit to human disease. Nature *529*, 326–335.

Wang, Q., Zhang, Z., Blackwell, K., and Carmichael, G.G. (2005). Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin. Curr. Biol. CB *15*, 384–391.

Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

Waudby, C.A., Dobson, C.M., and Christodoulou, J. (2019). Nature and Regulation of Protein Folding on the Ribosome. Trends Biochem. Sci. *44*, 914–926.

Weber, V., Wernitznig, A., Hager, G., Harata, M., Frank, P., and Wintersberger, U. (1997). Purification and nucleic-acid-binding properties of a Saccharomyces cerevisiae protein involved in the control of ploidy. Eur. J. Biochem. *249*, 309–317.

Woo, H.-H., Yi, X., Lamb, T., Menzl, I., Baker, T., Shapiro, D.J., and Chambers, S.K. (2011). Posttranscriptional suppression of proto-oncogene c-fms expression by vigilin in breast cancer. Mol. Cell. Biol. *31*, 215–225.

Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. Wiley Interdiscip. Rev. RNA *8*.

Yang, W.L., Wei, L., Huang, W.Q., Li, R., Shen, W.Y., Liu, J.Y., Xu, J.M., Li, B., and Qin, Y. (2014). Vigilin is overexpressed in hepatocellular carcinoma and is required for HCC cell proliferation and tumor growth. Oncol. Rep. *31*, 2328–2334.

Youn, J.-Y., Dunham, W.H., Hong, S.J., Knight, J.D.R., Bashkurov, M., Chen, G.I., Bagci, H., Rathod, B., MacLeod, G., Eng, S.W.M., et al. (2018). High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. Mol. Cell *69*, 517–532.e11.

Zhang, X., and Shan, S. (2014). Fidelity of cotranslational protein targeting by the signal recognition particle. Annu. Rev. Biophys. *43*, 381–408.

Zhou, J., Wang, Q., Chen, L.-L., and Carmichael, G.G. (2008). On the mechanism of induction of heterochromatin by the RNA-binding protein vigilin. RNA N. Y. N *14*, 1773–1781.

SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst habe und sämtliche Quellen, einschließlich Internetquellen, die unverändert oder abgewandelt wiedergegeben werden, insbesondere Quellen für Texte, Grafiken, Tabellen und Bilder, als solche kenntlich gemacht habe.

Ich versichere, dass ich die vorliegende Abschlussarbeit noch nicht für andere Prüfungen eingereicht habe.

Mir ist bekannt, dass bei Verstößen gegen diese Grundsätze ein Verfahren wegen Täuschungsversuchs bzw. Täuschung gemäß der fachspezifischen Prüfungsordnung und/oder der Fächerübergreifenden Satzung zur Regelung von Zulassung, Studium und Prüfung der Humboldt-Universität zu Berlin (ZSP-HU) eingeleitet wird.

Datum ..... Unterschrift .....

APPENDIX

Abbreviations

3'UTR | three prime untranslated region
4sU | 4-thiouridine
5'UTR | five prime untranslated region
BirA | Biotin ligase
bp | base pairs
C | cytosine
cDNA | complementary DNA
CDS | coding sequence
CRISPR | clustered regularly interspaced short palindromic repeats
C-terminus | carboxyl-terminus
DMEM | Dulbeco´s modified Eagle´s medium
DNA | deoxyribonucleic acid
EDTA | ethylenediaminetetraacetic acid
ER | endoplasmic reticulum
FBS | fetal bovine serum
FISH | fluorescent in situ hybridization
Fluc | Firefly luciferase
gRNA 1/2 | guide RNA 1/2
Gluc | *Gaussia* luciferase
GO | gene ontology
HDLBP | high-density lipoprotein binding protein
HEK | human embryonic kidney
HRP | horseradish peroxidase
iBAQ | intensity-based absolute quantification
irCLIP | infrared crosslinking and immunoprecipitation
KH | hnRNP K homology
KO | knockout
MFI | mean fluorescent intensity
mRNA | messenger RNA
nt | nucleotide
N-terminus | amino-terminus
PAR-CLIP | photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
PBS | phosphate buffered saline
PCR | polymerase chain reaction
PIC | pre-initiation complex
pSILAC | pulsed stable isotype labelling with amino acids in cell culture
RBP | RNA-binding protein
RIP | RNA immunoprecipitation
RNA | ribonucleic acid
rpm | revolutions per minute
RREs | RNA recognition elements
rRNA | ribosomal RNA
S | sedimentation coefficient
SDS | sodium dodecyl sulfate
SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEAP | secreted alkaline phosphatase
SECReTE | secretion enhancing *cis*-regulatory targeting element
SignalP/SP | signal peptide
SRP | signal recognition particle
TBST | tris buffered saline with Tween
TCA | trichloroacetic acid
TM | transmembrane domain
TPM | transcripts per million
tRNA | transfer RNA
TS | targeting signal
U | uracil
UV | ultraviolet
WT | wild-type

List of figures

Introduction

Results

Publications and presentations

Publications

- Stoehr, R., Taubert, H., **Zinnall, U.**, Giedl, J., Gaisa, N.T., Burger, M., Ruemmele, P., Hurst, C.D., Knowles, M.A., Wullich, B., Hartmann, A., **2015**. Frequency of TERT Promoter Mutations in Prostate Cancer. Pathobiol. J. Immunopathol. Mol. Cell. Biol. 82, 53–57.
- Giedl, J., Rogler, A., Wild, A., Riener, M.-O., Filbeck, T., Burger, M., Rümmele, P., Hurst, C., Knowles, M., Hartmann, A., **Zinnall, U.**, Stoehr, R., **2016**. TERT Core Promotor Mutations in Early-Onset Bladder Cancer. J. Cancer 7, 915–920.
- Milek, M., Imami, K., Mukherjee, N., Bortoli, F.D., **Zinnall, U.**, Hazapis, O., Trahan, C., Oeffinger, M., Heyd, F., Ohler, U., Selbach, M., Landthaler, M., **2017**. DDX54 regulates transcriptome dynamics during DNA damage response. Genome Res. 27, 1344–1359.
- **Zinnall, U.**, Weyerer, V., Compérat, E., Camparo, P., Gaisa, N.T., Knuechel-Clarke, R., Perren, A., Lugli, A., Toma, M., Baretton, G., Kristiansen, G., Wirtz, R.M., Cheng, L., Wullich, B., Stoehr, R., Hartmann, A., Bertz, S., **2018**. Micropapillary urothelial carcinoma: evaluation of HER2 status and immunohistochemical characterization of the molecular subtype. Hum. Pathol. 80, 55–64.
- Weyerer, V., Weisser, R., Moskalev, E.A., Haller, F., Stoehr, R., Eckstein, M., **Zinnall, U.**, Gaisa, N.T., Compérat, E., Perren, A., Keck, B., Allory, Y., Kristiansen, G., Wullich, B., Agaimy, A., Hartmann, A., Bertz, S., **2019**. Distinct genetic alterations and luminal molecular subtype in nested varaint of urothelial carcinoma. Histopathology. https://doi.org/10.1111/his.13958

Manuscripts in revision (as of October 2020)

- HDLBP discriminates between ER-targeted and cytosolic mRNAs to promote protein synthesis of transmembrane and secreted proteins
  U. Zinnall[+]; M. Milek[+]; C. H. Vieira-Vieira; S. Müller, G. Mastrobuoni; O. Hazapis; I. Minia; S. del Guidice; N. Bley; S. Kempa; S. Hüttelmaier; M. Selbach; M. Landthaler [Nature Communications]
  [+]These authors contributed equally
- The architecture of protein synthesis in the developing neocortex at near-atomic resolution reveals Ebp1-mediated neuronal proteostasis at the 60S tunnel exit
  M. L. Kraushar; F. Krupp; D. Harnett; P. Turko; M. C. Ambrozkiewicz; T. Sprink; K. Imami; M. Günnigmann; **U. Zinnall**; C. H. Vieira-Vieira; T. Schaub; A. Münster-Wandowski; J. Bürger; E. Borisova; H. Yamamoto; M. Rasin; U. Ohler; D. Beule; T. Mielke; V. Tarabykin; M. Landthaler; G. Kramer; I. Vida; M. Selbach; C. M. T. Spahn [Molecular Cell]
- 4EHP and GIGYF1/2 mediate translation-coupled messenger RNA decay
  R. Weber; M. Chung, C. Keskeny; **U. Zinnall**; M. Landthaler, E. Valkov, E. Izaurralde; C. Igreja [Cell Reports]

Presentations

- October 2016: Poster presentation at the EMBO | EMBL symposia “The Complex Life of mRNA” in Heidelberg
- July 2017: Poster presentation at the EMBO conference “RNA Localisation and Local Translation” in Italy
- October 2018: Poster presentation at the EMBO conference “Endoplasmic reticulum function in health and disease” in Italy
- June 2019: Poster presentation at “The RNA Localization and Local Translation Conference” in Snowmass Village, Colorado, US

ACKNOWLEDGEMENTS

During all these years I was wondering how this moment will feel like: writing the last missing piece of a whole 103 page long thesis. I am thinking back of a more than 5 year long journey with the mystery of only one single RNA-binding protein and so many people contributing to it… Everything started back in February 2015 with my interview in Markus office in Buch and a fox passing by his window. Strangely enough, despite my distraction, I ended up being hired. Thank you Markus, for the great opportunity to work in your lab and your support!

But Buch didn´t just had to offer foxes. I worked there the first time with RNA, learned how to swear in Polish and Italian, how to do PAR-CLIP and how to look forward to Spaghetti Bolognese every Friday in Mensa. Thank you, all my dear colleagues, for your help, patience and all the fun!!! I would like to especially acknowledge Roberto for his major contribution to the fun part. I will never forget his talent for all kinds of animal sounds. I would like to especially acknowledge Marcin and Karol for their contribution to the help part and for showing me step by step all kinds of experiments. Moreover, I would like to especially acknowledge Nico and Emanuel for their contribution to the patience part because they tried to teach me some bioinformatic skills...

Most importantly, I am incredibly thankful for the past ~1.5 years. Not only that I had the chance to work in Mitte and to discover how diverse lunch breaks can be, also my project took a tremendous twist. Before, mainly Orsalia and I tried to make sense out of the vast majority of data and Orsalia, thanks a lot for doing your best! But only after Miha joined everything started to make sense. Miha, thank you for everything, for supervising me at the beginning, for being always open to stupid questions and for all your expertise!

Next, I would like to extent my acknowledgments from our lab, to all associated lab members, to our whole floor back in Buch but also here in Mitte, to the whole building and the whole campus! I always enjoyed the atmosphere and I am very thankful for all the courses I could take, for the retreats, the conferences I could go to, the beer hours and all the collaborations that I was involved in! Thank you!

I would also like to thank my committee members Marina and Matthias for their feedback and Benedikt and Florian who agreed to review this thesis. And for taking the time to correct and discuss this thesis I would like to acknowledge Kasia, Igor, Nico, Emanuel and Miha. Thank you!

Last but not least, I always aimed to keep a healthy and happy work-life balance (as stupid as the term sounds) and that wouldn´t have been possible without my family and friends. I would like to especially acknowledge my parents and grandparents for always encouraging and believing in me particularly when I didn´t. Most importantly, I am truly grateful for the unlimited moral support of the best roommate I ever had, my friend and partner Joseph. Thank you all from the bottom of my ♥!